

# A VARIATION NORM CARLESON THEOREM

RICHARD OBERLIN   ANDREAS SEEGER   TERENCE TAO  
CHRISTOPH THIELE   JAMES WRIGHT

ABSTRACT. We strengthen the Carleson-Hunt theorem by proving  $L^p$  estimates for the  $r$ -variation of the partial sum operators for Fourier series and integrals, for  $p > \max\{r', 2\}$ . Four appendices are concerned with transference, a variation norm Menshov-Paley-Zygmund theorem, and applications to nonlinear Fourier transforms and ergodic theory.

## 1. INTRODUCTION

For an integrable function  $f$  on the circle group  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ , and  $k \in \mathbb{Z}$  we denote by  $\widehat{f}_k = \int_0^1 f(y) e^{-2\pi i k y} dy$  the Fourier coefficients and consider the partial sum operators  $S_n$  for the Fourier series,

$$(1) \quad S_n f(x) \equiv S[f](n, x) = \sum_{k=-n}^n \widehat{f}_k e^{2\pi i k x};$$

here  $n \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ . The celebrated theorem by Carleson [4] states that if  $f$  is square integrable then  $S_n f$  converges to  $f$  almost everywhere. Hunt [13] extended this result to  $L^p(\mathbb{T})$  functions, for  $1 < p < \infty$ , and proved the inequality

$$(2) \quad \left\| \sup_n |S_n f| \right\|_{L^p(\mathbb{T})} \leq C \|f\|_{L^p(\mathbb{T})}$$

for all  $f \in L^p(\mathbb{T})$ ; see also [10], [20], and [12] for other proofs of this fact.

The purpose of this paper is to strengthen the Carleson-Hunt result for  $L^p$  functions,  $1 < p < \infty$ , and show that, for  $r > \max\{2, p'\}$ , the (strong)  $r$ -variation of the sequence  $\{S_n f(x)\}_{n \in \mathbb{N}_0}$  is finite for almost every  $x \in [0, 1]$ . This can be interpreted as a statement about the rate of convergence. To fix notation, we consider real or complex valued sequences  $\{a_n\}_{n \in \mathbb{N}_0}$  and define their  $r$ -variation to be

$$(3) \quad \|a\|_{V^r} = \sup_K \sup_{n_0 < \dots < n_K} \left( \sum_{\ell=1}^K |a_{n_\ell} - a_{n_{\ell-1}}|^r \right)^{1/r}$$

where the sup is taken over all  $K$  and then over all increasing sequences of nonnegative integers  $n_0 < \dots < n_K$ . Note that the variation norms are monotone decreasing in the parameter  $r$ . Next, for a sequence  $F = \{F_n\}$  of Lebesgue measurable functions one defines the  $r$ -variation of  $F$  at  $x$ , sometimes denoted by  $\mathcal{V}^r F(x)$  as the  $V^r$  norm of the

---

*Date:* August 26, 2010.

1991 *Mathematics Subject Classification.* 42B15.

R.O. partially supported by NSF VIGRE grant DMS 0502315. A.S. partially supported by NSF grant DMS 0652890. T.T. partially supported by NSF Research Award DMS-0649473, the NSF Waterman award and a grant from the McArthur Foundation. C.Th. partially supported by NSF grant DMS 0701302. J.W. partially supported by an EPSRC grant.

sequence  $\{F_n(x)\}$ . We denote the  $r$ -variation of the sequence  $F_n = S_n f$  by  $\mathcal{V}^r S[f]$ . The variation norms and the  $r$ -variation operator can be defined in a similar fashion, if the index set  $\mathbb{N}_0$  is replaced by another subset of  $\mathbb{R}$  (often  $\mathbb{R}^+$  or  $\mathbb{R}$  itself).

Let  $r' := r/(r-1)$ , the conjugate exponent of  $r$ .

**Theorem 1.1.** *Suppose  $r > 2$  and  $r' < p < \infty$ . Then, for every  $f \in L^p(\mathbb{T})$ ,*

$$(4) \quad \|S[f]\|_{L^p(V^r)} \leq C_{p,r} \|f\|_{L^p}.$$

*At the endpoint  $p = r'$  a restricted weak type result holds; namely, for any  $f \in L^{r',1}(\mathbb{T})$  the function  $\mathcal{V}^r S[f]$  belongs to  $L^{r',\infty}(\mathbb{T})$ .*

It is immediate that (4) for  $r < \infty$  implies a quantitative form of almost everywhere convergence of Fourier series, improving over the standard qualitative result utilizing the weaker  $r = \infty$  inequality and convergence on a dense subclass of functions.

As will be discussed in Section 2, the conditions on the exponents in (4) are sharp. Moreover, in the endpoint case  $p = r'$  the Lorentz space  $L^{r',\infty}$  cannot be replaced by a smaller Lorentz space.

By standard transference arguments (see Appendix A) Theorem 1.1 is implied by a result on the partial (inverse) Fourier integral of a Schwartz function  $f$  on  $\mathbb{R}$  is defined as

$$\mathcal{S}[f](\xi, x) = \int_{-\infty}^{\xi} \widehat{f}(\eta) e^{2\pi i \eta x} d\eta$$

where  $\widehat{f}(\eta) = \int f(y) e^{-2\pi i y \eta} dy$  defines the Fourier transform of  $f$ .

**Theorem 1.2.** *Suppose  $r > 2$ . Then  $\mathcal{S}$  extends to a bounded operator  $\mathcal{S} : L^p \rightarrow L^p(V^r)$  for  $r' < p < \infty$ . Moreover  $\mathcal{S}$  maps  $L^{r',1}$  boundedly to  $L^{r',\infty}(V^r)$ .*

Note that if in the above definition of the mixed  $L^p(V^r)$  spaces we interchange the order between integration in the  $x$  variable and taking the supremum over the choices of  $K$  and the points  $\xi_0$  to  $\xi_K$  so that these choices become independent of the variable  $x$ , then the estimates corresponding to Theorem 1.2 are weaker; they follow from a square function inequality of Rubio de Francia [30] for  $p \geq 2$ , see also [29] for a related endpoint result for  $p < 2$ , and [18] for a proof of Rubio de Francia's inequality which is closer to the methods of this paper.

While the concept of  $r$ -variation norm is at least as old as Wiener's 1920s paper on quadratic variation [34], variational estimates have been pioneered by D. Lépingle ([22]) who proved them for martingales. Simple proofs of Lépingle's result based on jump inequalities have been given by Pisier and Xu [27] and by Bourgain [1], and applications to other families of operators in harmonic analysis such as families of averages and singular integrals have been considered in [1], and the subsequent papers [14], [2], [15] (*cf.* the bibliography of [15] for more references). Bourgain [1] used variation norm estimates (or related oscillation estimates which are intermediate in difficulty between maximal and variation norm estimates) to prove pointwise convergence results without previous knowledge that pointwise convergence holds for a dense subclass of functions. Such dense subclasses of functions, while usually available in the setting of analysis on Euclidean space, are less abundant in the ergodic theory setting. In Appendix D we

demonstrate the use of Theorem 1.2 in the setting of Wiener-Wintner type theorems as developed in [19]. We note that the Carleson-Hunt theorem has previously been generalized by using other norms in place of the variation norm, see for example the use of oscillation norms in [19], and the  $M_2^*$  norms in [8], [9].

We are also motivated by the fact that variation norms are in certain situations more stable under nonlinear perturbation than supremum norms. For example one can deduce bounds for certain  $r$ -variational lengths of curves in Lie groups from the corresponding lengths of the “trace” of the curves in the corresponding Lie algebras, see Appendix C for definitions and details. What we have in mind is proving Carleson type theorems for nonlinear perturbations of the Fourier transform as discussed in [25], [26]. Unfortunately the naive approach fails and the ultimate goal remains unattained since we only know the correlation between lengths of the trace and the original curve for  $r < 2$ , while the variational Carleson theorem only holds for  $r > 2$ . Nonetheless, this method allows one to see that a variational version of the Christ-Kiselev theorem [6] follows from a variational Menshov-Paley-Zygmund theorem which we prove in Appendix B. The variational Carleson inequality can be viewed as an endpoint in this theory.

Our proof of Theorem 1.2 will follow the method of [20] as refined in [12]. Naturally one has to invoke variation norm results in the setting of individual trees, which is achieved by adapting D. Lépingle’s result ([22]) to the setting of a tree. The authors initially had a proof of the case  $p > 2$  and  $r > p$  of Theorem 1.2 more akin to [20], while improvement to  $r > 2$  for such  $p$  provided a stumbling block. This stumbling block was removed by better accounting for trees of given energy, as described in the remarks leading to Proposition 4.3. In Section 3 we reduce the problem to that of bounding certain model operators which map  $f$  to linear combinations of wave-packets associated to collections of multitiles. In Section 5 we bound the model operators when the collection of multitiles is of a certain type called a tree; this bound is in terms of two quantities, energy and density, which are associated to the tree. These quantities are defined in Section 4 and an algorithm is given to decompose an arbitrary collection of multitiles into a union of trees with controlled energy and density. Section 6 contains two auxiliary estimates. All these ingredients are combined to complete the proof in Section 7.

*Some notation.* For two quantities  $A$  and  $B$  let  $A \lesssim B$  denote the statement that  $A \leq CB$  for some constant  $C$  (possibly depending on the parameters  $p$  and  $r$ ). The Lebesgue measure of a set  $E$  is denoted either by  $|E|$  or by  $\text{meas}(E)$ . The indicator function of  $E$  is denoted by  $\mathbb{1}_E$ . For a subset  $E \subset \mathbb{R}$  and  $a \in \mathbb{R}$  we set  $a + E = E + a = \{x : x - a \in E\}$ . If  $I$  is a finite interval  $I$  with center  $c(I)$  we denote by  $CI$  the  $C$  dilate of  $I$  with respect to its center, i.e. the set of all  $x$  for which  $c(I) + \frac{x - c(I)}{C} \in I$ .

## 2. OPTIMALITY OF THE EXPONENTS

Since Theorem 1.2 implies Theorem 1.1 (*cf.* Appendix A) we have to discuss the optimality only for the Fourier series case. The necessity of the condition  $r > 2$  follows from a corresponding result for the Cesaro means; its proof by Jones and Wang [16] was based on a probabilistic result of Qian [28].

We show the necessity of the condition  $p > r'$  in Theorem 1.1. Let

$$D_n(x) = \sum_{k=-n}^n e^{2\pi i k x} = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)},$$

the Dirichlet kernel, and let  $f^N$  be the de la Vallée-Poussin kernel which is defined by  $f^N = 2K_{2N+1} - K_N$  via the Fejér kernel  $K_N = (N+1)^{-1} \sum_{j=0}^N D_j$ . Then  $[\widehat{f^N}]_k = 1$  for  $|k| \leq N+1$  and thus  $S_n f^N = D_n$  for  $|n| \leq N+1$ . We have  $\|f_N\|_{L^1(\mathbb{T})} = O(1)$ , and  $\|f_N\|_\infty = O(N)$  and therefore  $\|f_N\|_{L^{p,q}(\mathbb{T})} = O(N^{1-1/p})$ .

Let  $N \gg 10^3$  and  $8N^{-1} \leq x \leq 1/8$ . Let  $K = K(x)$  be the largest integer  $< Nx$ . Then for  $0 \leq k < 2K(x)$  there are integers  $n_k(x) \leq N$  so that  $(2n_k(x) + 1)x \in (\frac{1}{4} + k, \frac{3}{4} + k)$ , in particular  $n_k(x) < n_{k+1}(x)$  for  $k < 2K(x) - 1$ . Observe  $\sin((2n_{2j}(x) + 1)\pi x) > \sqrt{2}/2$  and  $\sin((2n_{2j+1}(x) + 1)\pi x) < -\sqrt{2}/2$  for  $0 \leq j \leq K(x) - 1$ . This gives

$$\begin{aligned} \left( \sum_{j=0}^{K(x)-1} |S_{n_{2j+1}(x)} f^N - S_{n_{2j}(x)} f^N|^r \right)^{1/r} &= \left( \sum_{j=0}^{K(x)-1} |D_{n_{2j+1}(x)} - D_{n_{2j}(x)}|^r \right)^{1/r} \\ &\geq \frac{K(x)^{1/r} \sqrt{2}}{\sin(\pi x)} \geq cN^{1/r} x^{1/r-1}, \end{aligned}$$

and this implies for large  $N$

$$\frac{\|\mathcal{V}^r(Sf^N)\|_{L^{p,s}}}{\|f_N\|_{L^{p,1}}} \geq c_{p,s} \begin{cases} N^{\frac{1}{p}-\frac{1}{r'}} & \text{if } p < r', \\ (\log N)^{1/s} & \text{if } p = r'. \end{cases}$$

Thus the  $L^p \rightarrow L^p(V^r)$  boundedness does not hold for  $p < r'$ ; moreover the  $L^{r',1} \rightarrow L^{r',s}(V^r)$  boundedness does not hold for  $s < \infty$ .

### 3. THE MODEL OPERATORS

We shall show in appendix A how to deduce Theorem 1.1 from Theorem 1.2. To start the proof of the main Theorem 1.2, we describe some reductions to model operators involving wave packet decompositions.

First, by interpolation it suffices to prove for  $p \geq r'$  the restricted weak type  $L^{p,1} \rightarrow L^{p,\infty}(V_r)$  bound. Next, by the monotone convergence theorem it suffices to estimate  $L^{p,\infty}(V_r)$  on finite  $x$ -intervals  $[-A, A]$ , with constant independent of  $A$ . By another application of the monotone convergence theorem it suffices, for any fixed  $K$ , to prove the  $L^{p,1} \rightarrow L^{p,\infty}([-A, A])$  bound for

$$(5) \quad \sup_{\xi_0 \leq \dots \leq \xi_K} \left( \sum_{\ell=1}^K |\mathcal{S}f(\xi_\ell, x) - \mathcal{S}f(\xi_{\ell-1}, x)|^r \right)^{1/r}$$

where the sup is taken over all  $(\xi_0, \dots, \xi_K)$  with  $\xi_{\ell-1} \leq \xi_\ell$  for  $\ell = 1, \dots, K$ . Moreover, by the density of Schwartz functions in  $L^{p,1}$  it suffices to prove a uniform estimate for all Schwartz functions. Note that for any Schwartz function  $f$  the expression  $\mathcal{S}[f](\xi, x)$  depends continuously on  $(\xi, x)$ . Therefore it suffices to bound the expression analogous to (5) where we impose the strict inequality  $\xi_{\ell-1} < \xi_\ell$  for  $\ell = 1, \dots, K$ . Moreover, by the continuity it suffices for each finite set  $\Xi \subset \mathbb{R}$  to prove bounds for this expression under

the assumption that the  $\xi_\ell$  belong to  $\Xi$ , and we may also assume that  $\Xi$  does not contain any numbers of the form  $n2^m$  with  $m, n \in \mathbb{Z}$  (*i.e.* no endpoints of dyadic intervals).

We may now linearize the variation norm. Fix  $K \in \mathbb{N}$ , measurable real valued functions  $\xi_0(x) < \dots < \xi_K(x)$ , with values in  $\Xi$  and measurable complex valued functions  $a_1(x), \dots, a_K(x)$  satisfying

$$|a_1(x)|^{r'} + \dots + |a_K(x)|^{r'} = 1.$$

Let

$$\mathcal{S}'[f](x) = \sum_{k=1}^K (\mathcal{S}[f](\xi_k(x), x) - \mathcal{S}[f](\xi_{k-1}(x), x)) a_k(x).$$

Theorem 1.2 will now follow from the estimate

$$(6) \quad \|\mathcal{S}'[f]\|_{L^{p,\infty}(\mathbb{R})} \leq C \|f\|_{L^{p,1}(\mathbb{R})}$$

where  $C$  is independent of  $K$ ,  $\Xi$  and the linearizing functions, and where  $f$  is any Schwartz function. Finally, in order to prove (6) for any fixed  $\Xi$  we may assume that  $\widehat{f}$  has compact support in  $\mathbb{R} \setminus \Xi$  since the space of Schwartz functions with this property is dense in  $L^{p,1}$ ,  $1 < p < \infty$ .

Let  $\mathcal{D} = \{[2^k m, 2^k(m+1)) : m, k \in \mathbb{Z}\}$  be the set of dyadic intervals. A *tile* will be any rectangle  $I \times \omega$  where  $I, \omega$  are dyadic intervals, and  $|I||\omega| = 1/2$ . We will write  $\mathcal{S}'$  as the sum of wave packets adapted to tiles, and then decompose the operator into a finite sum of model operators by sorting the wave packets into a finite number of classes. For each  $k$ ,

$$\mathcal{S}[f](\xi_k, x) - \mathcal{S}[f](\xi_{k-1}, x) = \int \mathbb{1}_{(\xi_{k-1}, \xi_k)}(\xi) \widehat{f}(\xi) e^{2\pi i \xi x} d\xi.$$

To suitably express the difference above as a sum of wave packets, we will first need to construct a partition of  $\mathbb{1}_{(\xi_{k-1}, \xi_k)}$  adapted to certain dyadic intervals. The fact that  $(\xi_{k-1}, \xi_k)$  has two boundary points instead of the one from  $(-\infty, \xi_k)$  will necessitate a slightly more involved discretization argument than that in [20].

For any  $\xi < \xi'$ , let  $\mathbf{J}_{\xi, \xi'}$  be the set of maximal dyadic intervals  $J$  such that  $J \subset (\xi, \xi')$  and  $\text{dist}(J, \xi) \geq |J|$ ,  $\text{dist}(J, \xi') \geq |J|$ . Let  $\nu$  be a  $C^\infty$  function from  $\mathbb{R}$  to  $[0, 1]$  which vanishes on  $(-\infty, -10^{-2}]$ , is identically equal to 1 on  $[10^{-2}, \infty)$ , and so that  $\nu'(x) \geq 0$  for  $-10^{-2} < x < 10^{-2}$ . Given an interval  $J = [a, b)$ , and  $i \in \{-1, 0, 1\}$ , define

$$\varphi_{J,i}(\xi) = \nu\left(\frac{\xi - a}{2^i(b-a)}\right) - \nu\left(\frac{\xi - b}{b-a}\right).$$

Thus if  $c(J) = \frac{a+b}{2}$ , the center of  $J$ , then

$$(7) \quad \varphi_{J,i}(\xi) = \nu_i\left(\frac{\xi - c(J)}{|J|}\right) \text{ where } \nu_i(\eta) = \nu(2^{-i}(\eta + \tfrac{1}{2})) - \nu(\eta - \tfrac{1}{2}),$$

and we notice for  $i \in \{-1, 0, 1\}$  both  $\nu_i$  and  $\sqrt{\nu_i}$  are  $C^\infty$  functions supported in  $[-\frac{13}{25}, \frac{13}{25}]$  (more precisely in  $[-\frac{13}{25}, \frac{51}{100}]$ ). Hence  $\varphi_{J,i}$  is supported on a  $\frac{26}{25}$ -dilate of  $J$  with respect to its center.

For each  $J \in \mathbf{J}_{\xi, \xi'}$ , one may check that there is a unique interval  $J' \in \mathbf{J}_{\xi, \xi'}$  which lies strictly to the left of  $J$  and satisfies  $\text{dist}(J', J) = 0$ , and one may check that  $J'$  has size

$|J|/2, |J|$ , or  $2|J|$ . We define  $\varphi_J = \varphi_{J, i(J)}$  where  $i(J)$  is chosen so that  $|J'| = 2^{i(J)}|J|$ . Then

$$(8) \quad \mathbb{1}_{(\xi, \xi')}(\eta) \widehat{f}(\eta) = \sum_{J \in \mathbf{J}_{\xi, \xi'}} \varphi_J(\eta) \widehat{f}(\eta).$$

Since we assume that  $\widehat{f}$  is compactly supported in  $\mathbb{R} \setminus \Xi$  we see that for every pair  $\xi < \xi'$  with  $\xi, \xi' \in \Xi$  only a finite number of dyadic intervals  $J \in \mathbf{J}_{\xi, \xi'}$  are relevant in (8).

We now write each multiplier  $\varphi_J$  as the sum of wave packets. For every tile  $P = I \times J$ , define

$$\phi_P(x) = |I|^{1/2} \mathcal{F}^{-1}[\sqrt{\varphi_J}](x - c(I))$$

where  $c(I)$  is the center of  $I$  and  $\mathcal{F}^{-1}$  denotes the inverse Fourier transform. For each  $J$ , we then have

$$(9) \quad \sum_{|I|=1/(2|J|)} \langle f, \phi_{I \times J} \rangle \widehat{\phi_{I \times J}} = \widehat{f} \varphi_J.$$

To see this we use a Fourier series expansion (cf. [33]). We first observe that  $\widehat{\phi_P}(\xi) = \sqrt{|I|} \sqrt{\varphi_J(\xi)} e^{-2\pi i c(I)\xi}$  and use  $\langle f, \phi_P \rangle = \langle \widehat{f}, \widehat{\phi_P} \rangle$ . Now let parametrize the centers of the dyadic intervals  $I$  of length  $L$  by  $-(k - \frac{1}{2})L$ ,  $k \in \mathbb{Z}$ . Set  $g_J(\omega) := [\sqrt{\varphi_J} \widehat{f}](c(J) + L^{-1}\omega) e^{\pi i \omega}$  and note that  $g_J$  is supported in  $[-\frac{13}{50}, \frac{13}{50}]$ . The left hand side of (9) is equal to

$$\begin{aligned} & \sqrt{\varphi_J(\xi)} \sum_k \int \widehat{f}(\eta) \sqrt{\varphi_J(\eta)} e^{-2\pi i (kL - \frac{L}{2})\eta} L d\eta e^{2\pi i (kL - \frac{L}{2})\xi} \\ &= \sqrt{\varphi_J(\xi)} e^{-\pi i L(\xi - c(J))} \sum_{k \in \mathbb{Z}} \int_{-1/2}^{1/2} g_J(\omega) e^{-2\pi i k \omega} d\omega e^{2\pi i k L(\xi - c(J))} \\ &= \sqrt{\varphi_J(\xi)} e^{-\pi i L(\xi - c(J))} g_J(L(\xi - c(J))) = \widehat{f}(\xi) \varphi_J(\xi) \end{aligned}$$

which gives (9). This in turn yields the representation of  $\mathcal{S}'[f]$  in terms of wave packets:

$$(10) \quad \mathcal{S}'[f](x) = \sum_{k=1}^K \left( \sum_{J \in \mathbf{J}_{\xi_{k-1}(x), \xi_k(x)}} \sum_{|I|=1/(2|J|)} \langle f, \phi_{I \times J} \rangle \phi_{I \times J} \right) a_k(x).$$

For the function  $f$  under consideration the above Fourier series expansion converges in  $L^2$ -Sobolev spaces of arbitrary high order and thus the convergence in (10) is uniform for  $x \in [-A, A]$ . Therefore it suffices, for any *finite* family  $\mathfrak{P}$  of tiles, to consider the operator  $\mathcal{S}''$  defined by

$$(11) \quad \mathcal{S}''[f](x) = \sum_{k=1}^K \left( \sum_{\substack{(I, J) \in \mathfrak{P} \\ J \in \mathbf{J}_{\xi_{k-1}(x), \xi_k(x)}}} \langle f, \phi_{I \times J} \rangle \phi_{I \times J}(x) \right) a_k(x).$$

The wave packets will be sorted into a finite number of classes, each well suited for further analysis. Sorting is accomplished by dividing every  $\mathbf{J}_{\xi, \xi'}$  into a finite number of disjoint sets. These sets will be indexed by a fixed subset of  $\{1, 2, 3\} \times \{1, 2, 3, 4\}^2 \times \{\text{left}, \text{right}\}$ . Specifically, for each  $(m, n, \text{side}) \in \{1, 2, 3, 4\}^2 \times \{\text{left}, \text{right}\}$ , we define

- $\mathbf{J}_{\xi, \xi', (1, m, n, \text{side})} = \{J \in \mathcal{D} : J \subset (\xi, \xi'), \xi \text{ is in the interval } J - (m+1)|J|, \xi' \text{ is in the interval } J + (n+1)|J|, \text{ and } J \text{ is the } \textit{side}\text{-child of its dyadic parent}\}.$

- $\mathbf{J}_{\xi, \xi', (2, m, n, \text{side})} = \{J \in \mathcal{D} : J \subset (\xi, \xi'), \xi \text{ is in the interval } J - (m+1)|J|, \text{dist}(\xi', J) \geq n|J|, \text{ and } J \text{ is the } \textit{side}\text{-child of its dyadic parent}\}.$
- $\mathbf{J}_{\xi, \xi', (3, m, n, \text{side})} = \{J \in \mathcal{D} : J \subset (\xi, \xi'), \text{dist}(\xi, J) > m|J|, \xi' \text{ is in the interval } J + (n+1)|J|, \text{ and } J \text{ is the } \textit{side}\text{-child of its dyadic parent}\}.$

We will choose  $R \subset \{1, 2, 3\} \times \{1, 2, 3, 4\}^2 \times \{\text{left}, \text{right}\}$  so that for each  $\xi, \xi'$ , the collection  $\{\mathbf{J}_{\xi, \xi', \rho}\}_{\rho \in R}$  is pairwise disjoint and  $\mathbf{J}_{\xi, \xi'} = \cup_{\rho \in R} \mathbf{J}_{\xi, \xi', \rho}$ . We will also assume that for each  $\rho \in R$  there is an  $i(\rho) \in \{-1, 0, 1\}$  such that  $|J'| = 2^{i(\rho)}|J|$  for every  $\xi < \xi'$ ,  $J \in \mathbf{J}_{\xi, \xi', \rho}$  and  $J' \in \mathbf{J}_{\xi, \xi'}$  with  $J'$  strictly to the left of  $J$  and  $\text{dist}(J, J') = 0$ . One may check that these conditions are satisfied, say, for

$$R = \{(1, 2, 1, \text{left}), (1, 2, 2, \text{left}), (1, 3, 1, \text{left}), \\ (1, 3, 2, \text{left}), (2, 1, 1, \text{left}), (2, 1, 1, \text{right}), (2, 2, 1, \text{right}), \\ (3, 4, 1, \text{left}), (3, 3, 1, \text{right}), (3, 4, 2, \text{left})\}.$$

It now follows that

$$\mathcal{S}''[f] = \sum_{\rho \in R} \mathcal{S}^\rho[f]$$

where

$$\mathcal{S}^\rho[f](x) = \sum_{k=1}^K \left( \sum_{\substack{(I \times J) \in \mathfrak{P} \\ J \in \mathbf{J}_{\xi_{k-1}(x), \xi_k(x), \rho}}} \langle f, \phi_{I \times J} \rangle \phi_{I \times J} \right) a_k(x).$$

It will be convenient to rewrite each operator  $\mathcal{S}^\rho$  in terms of *multitiles*. A multitle will be a subset of  $\mathbb{R}^2$  of the form  $I \times \omega$  where  $I \in \mathcal{D}$  and where  $\omega$  is the union of three intervals  $\omega_l, \omega_u, \omega_h$  in  $\mathcal{D}$ . For each  $\rho = (i, m, n, \text{side}) \in R$ , we consider a set of  $\rho$ -multitiles which is parameterized by  $\{(I, \omega_u) : I, \omega_u \in \mathcal{D}, |I||\omega_u| = 1/2, \text{ and } \omega_u \text{ is the } \textit{side}\text{-child of its parent}\}$ . Specifically, given  $\omega_u = [a, b]$

- If  $\rho = (1, m, n, \text{side})$  then  $\omega_l = \omega_u - (m+1)|\omega_u|$  and  $\omega_h = \omega_u + (n+1)|\omega_u|$ .
- If  $\rho = (2, m, n, \text{side})$  then  $\omega_l = \omega_u - (m+1)|\omega_u|$  and  $\omega_h = [a + (n+1)|\omega_u|, \infty)$ .
- If  $\rho = (3, m, n, \text{side})$  then  $\omega_l = (-\infty, b - (m+1)|\omega_u|)$  and  $\omega_h = \omega_u + (n+1)|\omega_u|$ .

For  $i = 1, 2, 3$  we shall say that  $\rho$  is an  $i$ -index if  $\rho = (i, m, n, \text{side})$ . For every  $\rho$ -multitle  $P$ , let  $a_P(x) = a_k(x)$  if  $k$  satisfies  $1 \leq k \leq K$  and  $\xi_{k-1}(x) \in \omega_l$  and  $\xi_k(x) \in \omega_h$  (such a  $k$  would clearly be unique), and  $a_P(x) = 0$  if there is no such  $k$ . Then,

$$\mathcal{S}^\rho[f](x) = \sum_{P \in \mathbf{P}_\rho} \langle f, \phi_P \rangle \phi_P(x) a_P(x)$$

where, for each  $\rho$ -multitle  $P$ ,  $\phi_P(x) = \sqrt{|I|} \mathcal{F}^{-1}[\sqrt{\varphi_{\omega_u, i(\rho)}}](x - c(I))$  and  $\mathbf{P}_\rho$  denotes the set of all  $\rho$ -multitiles for which  $I \times \omega_u$  belongs to  $\mathfrak{P}$ .

Inequality (6) and hence Theorem 1.2 will then follow after proving the bound

$$(12) \quad \|\mathcal{S}^\rho[f]\|_{L^{p, \infty}} \lesssim \|f\|_{L^{p, 1}}$$

for each  $\rho \in R$ . We shall only give the proof of this estimate for the case that  $\rho$  is a 1-index or  $\rho$  is a 2-index, and the case where  $\rho$  is a 3-index can be deduced by symmetry considerations. Indeed, if  $P = (I, \omega_u) := ([a, b], [c, d])$  and  $\tilde{P} := ([-b, a], [-d, c])$  then



$\langle f, \phi_P \rangle \phi_P = \langle f(-\cdot), \tilde{\phi}_{\tilde{P}} \rangle \tilde{\phi}_{\tilde{P}}$  where, in the definition of the  $\tilde{\phi}_{\tilde{P}}$  the function  $\nu_i$  in (7) is replaced with  $\nu_i(-\cdot)$  (both are supported in  $(-\frac{13}{25}, \frac{13}{25})$ ). Now reflection sends a half open interval  $[a, b)$  to a half open interval  $(-b, -a]$ , however this plays no role in our symmetry argument if, as we do, we assume that the set  $\Xi$  does not contain endpoints of dyadic intervals. We then see that the estimation of  $S^\rho[f]$  for  $\rho = (3, m, n, \text{side})$  is equivalent to the estimation of a  $S^{\tilde{\rho}}[f(-\cdot)]$  where the corresponding set  $\{P\}$  of multitiles is replaced with a set  $\{\tilde{P}\}$  of index  $\tilde{\rho} = (2, n, m, \text{opposite side})$  and the set  $\Xi$  is replaced with  $\{\xi : -\xi \in \Xi\}$ . Beginning with (14) both  $\nu_i$  or  $\nu_i(-\cdot)$  are allowed in the definition of the functions  $\varphi_{J,i}$  and  $\phi_P$ .

By the usual characterization of  $L^{p,1}$  as superpositions of functions bounded by characteristic functions it suffices to show that

$$\text{meas}(\{x : |S^\rho[f](x)| > \lambda\}) \leq C^p \lambda^{-p} |F|$$

where  $F \subset \mathbb{R}$ ,  $|F| > 0$ ,  $|f| \leq \mathbb{1}_F$ ,  $\lambda > 0$ ,  $2 < r < \infty$ , and  $r' \leq p < (1/2 - 1/r)^{-1}$ . This is accomplished by proving that for every measurable  $E \subset \mathbb{R}$ ,

$$(13) \quad \text{meas}\left(\{x \in E : |S^\rho[f](x)| > C(|F|/|E|)^{1/p}\}\right) \leq |E|/2.$$

Indeed, if we set  $E_\lambda = \{x : |S^\rho[f](x)| > \lambda\}$  then by the finiteness of the set of tiles under consideration the set  $E_\lambda$  has a priori finite measure. If  $|E_\lambda| \leq C^p \lambda^{-p} |F|$  then there is nothing to prove. If the opposite inequality  $|E_\lambda| > C^p \lambda^{-p} |F|$  were true then  $\lambda > C(|F|/|E_\lambda|)^{1/p}$  and inequality (13) applied to  $E = E_\lambda$  would yield that  $|E_\lambda| \leq |E_\lambda|/2$ , a contradiction.

We finally note that, after possibly rescaling, we may assume that  $1 \leq |E| \leq 2$  in (13). The next four sections will be devoted to the proof of inequality (13) in this case.

#### 4. ENERGY AND DENSITY

Recall that  $S^\rho[f](x) = \sum_P \langle f, \phi_P \rangle \phi_P a_P$  where  $P$  ranges over an arbitrary finite collection of  $\rho$ -multitiles,  $\rho$  is a 1 or 2-index. It is our goal to show (13) and for this and the next chapter we fix the function  $f$  with  $|f| \leq \mathbb{1}_F$  and the set  $E$ .

Fix  $1 \leq C_3 < C_2 < C_1$ , with  $C_2 \in \mathbb{N}$ , such that for every multitile  $P$ ,

$$\begin{aligned} \text{supp}(\widehat{\phi}_P) &\subset C_3 \omega_u \\ C_2 \omega_u \cap C_2 \omega_l &= \emptyset, \quad C_2 \omega_u \cap \omega_h = \emptyset, \\ C_2 \omega_l &\subset C_1 \omega_u, \quad C_2 \omega_u \subset C_1 \omega_l; \end{aligned}$$

recall that dilations of finite intervals are with respect to their center. One may check that the values  $C_3 = 11/10$ ,  $C_2 = 2$ , and  $C_1 = 12$  satisfy all these properties.

The wave packet is adapted to the multitile  $P$ . As  $\widehat{\phi}_P$  is compactly supported (in  $C_3 \omega_u$ ) the function  $\phi_P$  cannot have compact support, but as a replacement we have the following bounds involving

$$w_I(x) := \frac{1}{|I|} \left(1 + \frac{|x - c(I)|}{|I|}\right)^{-N}$$



for a fixed large  $N \gg 10$ ; namely

$$(14) \quad \left| \frac{d^n}{dx^n} (\exp(-2\pi i c(\omega_u) \cdot) \phi_P)(x) \right| \leq C'(n) |I|^{(1/2)-n} |w_I(x)|$$

for each  $n \geq 0$ .

We are working with a given finite set of  $\rho$ -multitiles  $\mathbf{P}$  (with  $\rho$  a 1- or 2-index) and we let  $M_o = M_o(\mathbf{P})$  be the smallest integer  $M$  for which all tiles are contained in the square  $[-2^M, 2^M]^2$ . Throughout this paper we fix

$$\Xi^{\text{top}} = \{ \eta : |\eta| \leq C_1 2^{M_o+10}, \eta = n 2^{-M_o-10}, \text{ for some } n \in \mathbb{Z} \}$$

as the set of admissible top-frequencies for trees, as in the following definition.

**Definition.** Consider a triple  $\mathcal{T} = (T_{\mathcal{T}}, I_{\mathcal{T}}, \xi_{\mathcal{T}})$ , with a set of multitiles  $T_{\mathcal{T}}$ , a dyadic interval  $I_{\mathcal{T}} \subset [-2^{M_o}, 2^{M_o})$  and a point  $\xi_{\mathcal{T}} \in \Xi^{\text{top}}$ . We say that  $\mathcal{T}$  is a tree if the following properties are satisfied.

(i)  $I \subset I_{\mathcal{T}}$  for all  $P = (I, \omega_u) \in T_{\mathcal{T}}$ .

(ii) If  $P = (I, \omega_u)$  and  $\omega_m$  denotes the convex hull of  $C_2 \omega_u \cup C_2 \omega_l$  then

$$\omega_{\mathcal{T}} := \left[ \xi_{\mathcal{T}} - \frac{C_2-1}{4|I_{\mathcal{T}}|}, \xi_{\mathcal{T}} + \frac{C_2-1}{4|I_{\mathcal{T}}|} \right)$$

is contained in  $\omega_m$ .

We refer to  $I_{\mathcal{T}}$  as the top interval of the tree, and to  $\xi_{\mathcal{T}}$  as the top frequency of the tree.

In order not to overload the notation we usually refer to the set  $T$  as “the” tree (keeping in mind that it carries additional information of a top frequency and a top interval), and we shall also use the notation  $I_T$ ,  $\xi_T$  and  $\omega_T$  in place of  $I_{\mathcal{T}}$ ,  $\xi_{\mathcal{T}}$  and  $\omega_{\mathcal{T}}$ . With this convention we also define

**Definition.**

(i) A tree  $(T, I_T, \xi_T)$  is  $l$ -overlapping if  $\xi_T \in C_2 \omega_l$  for every  $P \in T$ .

(ii) A tree  $(T, I_T, \xi_T)$  is  $l$ -lacunary if  $\xi_T \notin C_2 \omega_l$  for every  $P \in T$ .

Notice that the union of two trees with the top data  $I_T, \xi_T$  is again a tree with the same top data. Also, the union of two  $l$ -overlapping trees with the same top data is again an  $l$ -overlapping tree with the same top data.

We split our finite collection of multitiles into a bounded number of subcollections satisfying certain separation conditions (i.e. henceforth all multitiles will be assumed to belong to a fixed subcollection).

**Separation assumptions.**

$$(15) \quad \text{If } P, P' \text{ satisfy } |\omega'_u| < |\omega_u|, \text{ then } |\omega'_u| \leq \frac{C_2 - C_3}{2C_1} |\omega_u|.$$

$$(16) \quad \text{If } P, P' \text{ satisfy } C_1 \omega_u \cap C_1 \omega'_u \neq \emptyset \text{ and } |\omega_u| = |\omega'_u| \text{ then } \omega_u = \omega'_u.$$

As immediate but important consequence of the separation assumptions is the frequently used

**Observation 4.1.** *Let  $T$  be a tree satisfying the separation properties (15) and (16). Then the following properties hold.*

- (i) *If  $T$  is an  $l$ -overlapping tree,  $P, P' \in T$ , and  $|\omega'_u| < |\omega_u|$  then  $C_3\omega_u \cap C_2\omega'_u = \emptyset$ .*
- (ii) *If  $T$  is an  $l$ -lacunary tree,  $P, P' \in T$ , and  $|\omega'_u| < |\omega_u|$  then  $C_3\omega_l \cap C_3\omega'_l = \emptyset$ .*
- (iii) *If  $P, P' \in T$ ,  $P \neq P'$ , and  $|\omega_u| = |\omega'_u|$ , then  $I \cap I' = \emptyset$ .*

As in previous proofs of Carleson's theorem (in particular [20]) we shall split the set of multitiles into subsets with controllable *energy* and *density* associated to the function  $f$  and the set  $E$ , respectively. Here we work with the following definitions.

**Definition.** Fix  $f \in L^2(\mathbb{R})$  and a measurable set  $E \subset \mathbb{R}$ . Given any collection of multitiles  $\mathbf{P}$  we define

$$\text{energy}(\mathbf{P}) = \sup_T \sqrt{\frac{1}{|I_T|} \sum_{P \in T} |\langle f, \phi_P \rangle|^2}$$

where the sup ranges over all  $l$ -overlapping trees  $T \subset \mathbf{P}$ .

Given a measurable set  $E \subset \mathbb{R}$  we set

$$\text{density}(\mathbf{P}) = \sup_T \left( \frac{1}{|I_T|} \int_E \left( 1 + \frac{|x - c(I_T)|}{|I_T|} \right)^{-4} \sum_{k=1}^K |a_k(x)|^{r'} \mathbb{1}_{\omega_T}(\xi_{k-1}(x)) dx \right)^{1/r'}$$

where the sup is over all non-empty trees  $T \subset \mathbf{P}$ .

*Remark.* Concerning the terminology, one can argue that the squareroot should be omitted in the definition of an energy. However we work with the above definition to conform to [20] and other papers in time-frequency analysis.

**Lemma 4.2.** *Let  $|f|$  be bounded by  $\mathbb{1}_F$ . For any family  $\mathbf{P}$  of multitiles the density of  $\mathbf{P}$  (with respect to the set  $E$ ) and the energy (with respect to  $f$ ) are bounded by a universal constant.*

*Proof.* Clearly the density is bounded by  $\int_{\mathbb{R}} (1 + |x|)^{-4} dx < 3$ . Concerning the energy bound we let  $T$  be any  $l$ -overlapping tree, and split  $f = f' + f''$  where  $f' = \mathbb{1}_{3I_T} f$ . We estimate  $\sum_{P \in T} |\langle f', \phi_P \rangle|^2 \leq \|f'\|_{L^2}^2 (\sum_{P \in T} |\langle f', \phi_P \rangle|^2)^{1/2}$ , by the Cauchy-Schwarz inequality. Now use that the supports of the  $\hat{\phi}_P$  are disjoint for different sizes of frequency intervals, and then, for a fixed size, use the bounds (14) (for  $n = 0$ ) to see that

$$\left\| \sum_{P \in T} \langle f', \phi_P \rangle \phi_P \right\|_{L^2} \leq \left( \sum_{j \in \mathbb{Z}} \left\| \sum_{\substack{P \in T \\ |\omega_u| = 2^{-j}}} \langle f', \phi_P \rangle \phi_P \right\|_{L^2}^2 \right)^{1/2} \lesssim \left( \sum_{P \in T} |\langle f', \phi_P \rangle|^2 \right)^{1/2}.$$

Hence,

$$\sum_{P \in T} |\langle f', \phi_P \rangle|^2 \lesssim \|f'\|_{L^2}^2 \leq |F \cap 3I_T| \lesssim |I_T|.$$

Furthermore, since  $|f''| \leq \mathbb{1}_{\mathbb{R} \setminus 3I_T}$ , we have the estimate

$$|\langle f'', \phi_P \rangle| \lesssim |I|^{1/2} (1 + \text{dist}(I, \mathbb{R} \setminus 3I_T)/|I|)^{-(N-1)} \lesssim |I|^{1/2} (|I|/|I_T|)^{N-1}$$

Summing in  $P$ , we obtain

$$\sum_{P \in T} |\langle f'', \phi_P \rangle|^2 \lesssim |I_T|.$$

Combining the estimates for  $f'$  and  $f''$  we see that  $\sum_{P \in T} |\langle f, \phi_P \rangle|^2 \lesssim |I_T|$  for every  $l$ -overlapping tree and it follows that the energy of  $\mathbf{P}$  with respect to  $f$  is bounded above by a universal constant.  $\square$

The following proposition allows one to decompose an arbitrary collection of multitiles into the union of trees, where the trees are divided into collections  $\mathbf{T}_j$  with the energy of trees from  $\mathbf{T}_j$  bounded by  $2^{-j}$ . The control over energy is balanced by an  $L^q$  bound for the functions  $N_{j,\ell} := \sum_{T \in \mathbf{T}_j} \mathbb{1}_{2^\ell I_T}$ . In contrast to [20] and [12], it is necessary here to consider  $q > 1$  and  $\ell > 0$  in order to effectively use the tree estimate Proposition 5.1 with  $q > 1$ . Note that such  $L^q$  bound for  $N_{j,\ell}$  is established by combining (17) and (18) below. The bound (20) permits one to make further decompositions to take advantage of large  $|F|$  in the  $L^q$  bound for the  $N_{j,\ell}$  while maintaining compatibility with bounds for trees with a fixed density obtained from Proposition 4.4.

**Proposition 4.3.** *Let  $\mathcal{E} > 0$ , let  $|f|$  be bounded above by  $\mathbb{1}_F$  and let  $\mathbf{P}$  be a collection of multitiles with energy bounded above by  $\mathcal{E}$ . Then, there is a collection of trees  $\mathbf{T}$  such that*

$$(17) \quad \sum_{T \in \mathbf{T}} |I_T| \lesssim \mathcal{E}^{-2} |F|$$

and

$$\text{energy} \left( \mathbf{P} \setminus \bigcup_{T \in \mathbf{T}} T \right) \leq \mathcal{E}/2,$$

and such that, for every integer  $\ell \geq 0$ ,

$$(18) \quad \left\| \sum_{T \in \mathbf{T}} \mathbb{1}_{2^\ell I_T} \right\|_{BMO} \lesssim 2^{2\ell} \mathcal{E}^{-2}.$$

Furthermore, if for some collection of trees  $\mathbf{T}'$ ,

$$(19) \quad \mathbf{P} = \bigcup_{T' \in \mathbf{T}'} T'$$

then

$$(20) \quad \sum_{T \in \mathbf{T}} |I_T| \lesssim \sum_{T' \in \mathbf{T}'} |I_{T'}|.$$

Above, and subsequently,  $\|\cdot\|_{BMO}$  denotes the dyadic  $BMO$  norm.

*Proof.* We select trees through an iterative procedure. First, if  $\text{energy}(\mathbf{P}) \leq \mathcal{E}/2$  then no tree is chosen and  $\mathbf{T} = \emptyset$ .

If  $\text{energy}(\mathbf{P}) \geq \mathcal{E}/2$  then we observe that there is an  $l$ -overlapping tree  $S \subset \mathbf{P}$  for which

$$(21) \quad \frac{1}{|I_S|} \sum_{P \in S} |\langle f, \phi_P \rangle|^2 \geq \mathcal{E}^2/4.$$

There is only a finite number of such trees and as  $S_1$  we choose one for which the top datum  $\xi_S$  is maximal (in  $\mathbb{R}$ ). Note that the maximality can be achieved as we restrict all top frequency data to the finite set  $\Xi^{\text{top}}$ . Let  $T_1$  be the tree in  $\mathbf{P}$  which has top data  $(\xi_{S_1}, I_{S_1})$  and which is maximal with respect to inclusion.

Suppose that trees  $S_k, T_k$  have been chosen for  $k = 1, \dots, j$ . Set

$$\mathbf{P}_j = \mathbf{P} \setminus \bigcup_{k=1}^j T_k$$

If  $\text{energy}(\mathbf{P}_j) \leq \mathcal{E}/2$  then we terminate the procedure, set  $\mathbf{T} = \{T_k\}_{1 \leq k \leq j}$  and  $n = j$ . Otherwise, we may find an  $l$ -overlapping tree  $S \subset \mathbf{P}_j$  such that (21) holds. Among  $l$ -overlapping in  $\mathbf{P}_j$  satisfying (21) choose one with maximal top-frequency (in  $\Xi^{\text{top}}$ ) and label this tree  $S_{j+1}$ . Let  $T_{j+1}$  be the maximal tree in  $\mathbf{P}_j$  which has with top data  $(\xi_{S_{j+1}}, I_{S_{j+1}})$  and which is maximal with respect to inclusion. This process will eventually stop since each  $T_j$  is nonempty and  $\mathbf{P}$  is finite.

*Proof of (17).* It suffices to show

$$(22) \quad \left( \frac{\mathcal{E}^2}{|F|} \sum_{j=1}^n |I_{S_j}| \right)^2 \lesssim \frac{\mathcal{E}^2}{|F|} \sum_{j=1}^n |I_{S_j}|.$$

Since the  $S_j$  satisfy (21), we have

$$\left( \frac{\mathcal{E}^2}{|F|} \sum_{j=1}^n |I_{S_j}| \right)^2 \leq 16|F|^{-2} \left( \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \right)^2.$$

Now

$$\begin{aligned} \left( \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \right)^2 &= \left( \left\langle \sum_{j=1}^n \sum_{P \in S_j} \langle f, \phi_P \rangle \phi_P, f \right\rangle \right)^2 \\ &\leq \|f\|_2^2 \left\| \sum_{j=1}^n \sum_{P \in S_j} \langle f, \phi_P \rangle \phi_P \right\|_2^2 \leq |F| \sum_{j=1}^n \sum_{P \in S_j} \sum_{k=1}^n \sum_{P' \in S_k} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \end{aligned}$$

where in the last inequality we used  $|f| \leq \mathbb{1}_F$ . By symmetry, it remains, for (22) to show that

$$(23) \quad \sum_{j=1}^n \sum_{k=1}^n \sum_{P \in S_j} \sum_{P' \in S_k: |I'|=|I|} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \lesssim \mathcal{E}^2 \sum_{j=1}^n |I_{S_j}|$$

and

$$(24) \quad \sum_{j=1}^n \sum_{k=1}^n \sum_{P \in S_j} \sum_{P' \in S_k: |I'| < |I|} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \lesssim \mathcal{E}^2 \sum_{j=1}^n |I_{S_j}|.$$

In both cases, we will use the estimate

$$(25) \quad |\langle \phi_P, \phi_{P'} \rangle| \lesssim \left( \frac{|I|}{|I'|} \right)^{1/2} \langle w_I, \mathbb{1}_{I'} \rangle.$$

which holds whenever  $|I'| \leq |I|$ .

Estimating the product of two terms by the square of their maximum, we see that the left side of (23) is

$$\leq 2 \sum_{j=1}^n \sum_{k=1}^n \sum_{P \in S_j} \sum_{P' \in S_k: |I'|=|I|} |\langle f, \phi_P \rangle|^2 |\langle \phi_P, \phi_{P'} \rangle|.$$

Recall that  $\langle \phi_P, \phi_{P'} \rangle = 0$  unless  $C_3\omega_u \cap C_3\omega'_u \neq \emptyset$ . Thus, by (16), (25) and the fact that the  $S_k$  are pairwise disjoint, we can estimate the last display by

$$2 \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \sum_{I': |I'|=|I|} \langle w_I, \mathbb{1}_{I'} \rangle \lesssim \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \lesssim \sum_{j=1}^n \mathcal{E}^2 |I_{S_j}|.$$

This finishes the proof of (23).

Applying the Cauchy-Schwarz inequality, we see that the left side of (24) is bounded by

$$\sum_{j=1}^n \left( \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \right)^{1/2} \left( \sum_{P \in S_j} \left( \sum_{k=1}^n \sum_{P' \in S_k: |I'| < |I|} |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \right)^2 \right)^{1/2}.$$

Twice using the fact that the energy of  $\mathbf{P}$  is bounded by  $\mathcal{E}$ , we see that the last display is

$$\leq \mathcal{E}^2 \sum_{j=1}^n |I_{S_j}|^{1/2} \left( \sum_{P \in S_j} \left( \sum_{k=1}^n \sum_{P' \in S_k: |I'| < |I|} |\langle \phi_P, |I_{P'}|^{1/2} \phi_{P'} \rangle| \right)^2 \right)^{1/2}.$$

Thus, to prove (24) it remains to show that, for each  $j$ ,

$$(26) \quad \sum_{P \in S_j} \left( \sum_{k=1}^n \sum_{P' \in S_k: |I'| < |I|} |\langle \phi_P, |I_{P'}|^{1/2} \phi_{P'} \rangle| \right)^2 \lesssim |I_{S_j}|.$$

Again, we only have  $|\langle \phi_P, |I_{P'}|^{1/2} \phi_{P'} \rangle|$  nonzero when  $C_3\omega_u \cap C_3\omega'_u \neq \emptyset$  which can only happen if  $\sup C_3\omega_u \in C_3\omega'_u$  or  $\inf C_3\omega_u \in C_3\omega'_u$ . Applying (25), we thus see that the left side of (26) is dominated by a constant times the expression

$$\sum_{P \in S_j} |I_P| \left( \sum_{k=1}^n \sum_{\substack{P' \in S_k: |I'| < |I| \\ \sup C_3\omega_u \in C_3\omega'_u}} \langle w_I, \mathbb{1}_{I'} \rangle \right)^2 + \sum_{P \in S_j} |I_P| \left( \sum_{k=1}^n \sum_{\substack{P' \in S_k: |I'| < |I| \\ \inf C_3\omega_u \in C_3\omega'_u}} \langle w_I, \mathbb{1}_{I'} \rangle \right)^2.$$

We now claim that, for each  $P \in S_j$  (with time interval  $I$ ),

$$(27) \quad \left( \sum_{k=1}^n \sum_{\substack{P' \in S_k: |I'| < |I| \\ \sup C_3\omega_u \in C_3\omega'_u}} \langle w_I, \mathbb{1}_{I'} \rangle \right)^2 \leq \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle$$

and that the same inequality with  $\sup$  replaced by  $\inf$  in the  $P'$  summation holds as well. To see this consider two multitiles  $P^1 = (I^1, \omega_u^1) \in S_{\kappa_1}$  and  $P^2 = (I^2, \omega_u^2) \in S_{\kappa_2}$ ,  $P^1 \neq P^2$ . so that  $|I^2| \leq |I^1|$  and  $C_3\omega_u^1 \cap C_3\omega_u^2 \neq \emptyset$ . The last condition implies  $\kappa_1 \neq \kappa_2$  (since  $S_{\kappa_1}$  and  $S_{\kappa_2}$  are  $l$ -overlapping). The inequality (27) is immediate if we can show that  $I^1$  and  $I^2$  are disjoint and if in addition  $|I^2| < |I^1|$ , then  $I_2$  does not belong to the top interval of the tree  $S_{\kappa_1}$ . Now, if  $|I^1| = |I^2|$ , then from (16) it follows that  $\omega_u^1 = \omega_u^2$  and hence, since  $P^1 \neq P^2$ , we have  $I^1 \cap I^2 = \emptyset$ . If  $|I^2| < |I^1|$ , then by (15)  $|\omega_u^1| \leq \frac{C_2 - C_3}{2C_1} |\omega_u^2|$ ; since  $C_3\omega_u^1 \cap C_3\omega_u^2 \neq \emptyset$  this implies that  $\inf C_2\omega_l^1 > \sup C_2\omega_l^2$ . As both trees are  $l$ -overlapping the top frequency of  $S_{\kappa_1}$  belongs to  $C_2\omega_l^1$  and is above the

top frequency of  $S_{\kappa_2}$  which belongs to  $C_2\omega_l^2$ . Thus by the maximality condition on the top frequency in the selection process of the trees we see that the tree  $S_{\kappa_1}$  was selected before the tree  $S_{\kappa_2}$ , i.e.  $\kappa_1 < \kappa_2$ . This implies that  $P_2$  does not belong to the tree  $T_{\kappa_1}$ , and since the interval  $\omega_{S_{\kappa_1}}$  is contained in the convex hull of  $\omega_l^2$  and  $\omega_u^2$  we see that the time intervals  $I^2$  and  $I_{T_{\kappa_1}} = I_{S_{\kappa_1}}$  cannot intersect. Thus  $I^1 \cap I^2 = \emptyset$  as  $I^1 \subset I_{T_{\kappa_1}}$ . This concludes the argument for (27).

Now by the disjointness condition we see that indeed the left hand side of (26) is bounded by a constant times

$$\begin{aligned} \sum_{P \in S_j} |I_P| \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle^2 &\lesssim \sum_{\ell: 2^\ell \leq |I_{S_j}|} 2^\ell \sum_{P \in S_j: |I|=2^\ell} \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle \\ &\lesssim |I_{S_j}| \sup_{\ell: 2^\ell \leq |I_{S_j}|} \sum_{P \in S_j: |I|=2^\ell} \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle \end{aligned}$$

One may check that, for each  $\ell$ ,

$$\sum_{P \in S_j: |I|=2^\ell} \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle \lesssim 1$$

and (26) follows. We have already seen that (26) implies (24); this completes the proof of (17).

*Proof of (18).* We need to show that for each dyadic interval  $J$ , we have

$$\frac{1}{|J|} \int_J \left| \sum_{T \in \mathbf{T}} \mathbb{1}_{2^\ell I_T}(x) - \frac{1}{|J|} \int_J \sum_{T \in \mathbf{T}} \mathbb{1}_{2^\ell I_T}(y) dy \right| dx \lesssim 2^{2\ell} \mathcal{E}^{-2}.$$

This is an immediate consequence of

$$(28) \quad \sum_{\substack{T \in \mathbf{T} \\ J \cap 2^\ell I_T \neq \emptyset, J}} |I_T| \lesssim \mathcal{E}^{-2} 2^\ell |J|.$$

Let

$$\tilde{\mathbf{T}} = \{T \in \mathbf{T} : I_T \subset 2^{\ell+1}J, |I_T| \leq |J|\}$$

and note that if  $T \in \mathbf{T}$  with  $2^\ell I_T \cap J \neq \emptyset, J$  then  $T \in \tilde{\mathbf{T}}$ . Write  $f = f' + f''$  where  $|f'| \leq \mathbb{1}_{F \cap 2^{\ell+5}J}$  and  $|f''| \leq \mathbb{1}_{F \cap \mathbb{R} \setminus 2^{\ell+5}J}$ .

We will write  $\tilde{\mathbf{T}}$  as the union of collections of trees  $\mathbf{T}^{\text{main}} \cup \mathbf{T}^0 \cup \mathbf{T}^1 \cup \dots$  each of which will have certain properties related to the energy. For each tree  $T \in \tilde{\mathbf{T}}$  there is an  $l$ -overlapping tree  $S = S(T)$  chosen in the algorithm above with  $I_S = I_T$  and

$$(29) \quad \frac{1}{|I_S|} \sum_{P \in S} |\langle f, \phi_P \rangle|^2 \geq \mathcal{E}^2/4.$$

Let

$$\mathbf{T}^0 = \{T \in \tilde{\mathbf{T}} : \frac{1}{|I_S|} \sum_{P \in S} |\langle f'', \phi_P \rangle|^2 \geq \mathcal{E}^2/16\}.$$

For  $j \geq 1$ , define

$$\mathbf{T}^j = \{T \in \tilde{\mathbf{T}} : \sup_{\substack{S' \subset S(T) \\ |I_{S'}|=2^{-j}|I_T|}} \frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f'', \phi_P \rangle|^2 \geq \mathcal{E}^2/16\}$$

where, for each  $T$ , the sup above is taken over all  $l$ -overlapping trees  $S'$  with  $S' \subset S(T)$ . Finally, let

$$\mathbf{T}^{\text{main}} = \{T \in \tilde{\mathbf{T}} \setminus (\mathbf{T}^0 \cup \mathbf{T}^1 \cup \dots)\}.$$

We split the sum (28) into the “main” term involving trees in  $\mathbf{T}^{\text{main}}$  and an error term involving  $\cup_{j \geq 0} \mathbf{T}_j$ .

We first consider the main term. It is our objective to prove

$$(30) \quad \sum_{T \in \mathbf{T}^{\text{main}}} |I_T| \lesssim \mathcal{E}^{-2} 2^\ell |J|.$$

Let  $T \in \mathbf{T}^{\text{main}}$  and let  $S'$  be any  $l$ -overlapping tree contained in  $S$  satisfying  $|I_{S'}| \leq |I_S|$ . Since the energy of  $\mathbf{P}$  is bounded by  $\mathcal{E}$  and since  $T$  is not in any  $\mathbf{T}^j$ , we have

$$\frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f', \phi_P \rangle|^2 \leq 2 \frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f, \phi_P \rangle|^2 + 2 \frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f'', \phi_P \rangle|^2 \lesssim \mathcal{E}^2.$$

From (21) and the fact that  $T \notin \mathbf{T}^0$ , we have

$$\frac{1}{|S(T)|} \sum_{P \in S(T)} |\langle f', \phi_P \rangle|^2 \geq \mathcal{E}^2/8 - \mathcal{E}^2/16 = \mathcal{E}^2/16.$$

This inequality allows us to essentially repeat the above proof of (17). The  $l$ -overlapping trees  $S(T)$  form a (finite or infinite) subsequence of the sequence  $S_j$  which we denote by  $S_{j(\nu)}$  so that we have  $j(\nu) > j(\nu')$  for  $\nu > \nu'$ . We need to prove the analogue of (22) which is

$$(31) \quad \frac{\mathcal{E}^2}{|F \cap 2^{\ell+5} J|} \sum_{\nu=1}^n |I_{S_{j(\nu)}}| \lesssim 1$$

and as before we are aiming to estimate the square of the expression on the left hand side by the expression itself. The  $S_{j(\nu)}$  satisfy

$$\frac{1}{|S_{j(\nu)}|} \sum_{P \in S_{j(\nu)}} |\langle f', \phi_P \rangle|^2 \geq \frac{\mathcal{E}^2}{16}$$

and therefore

$$\left( \frac{\mathcal{E}^2}{|F \cap 2^{\ell+5} J|} \sum_{\nu=1}^n |I_{S_{j(\nu)}}| \right)^2 \leq 256 |F \cap 2^{\ell+5} J|^{-2} \left( \sum_{\nu=1}^n \sum_{P \in S_{j(\nu)}} |\langle f, \phi_P \rangle|^2 \right)^2.$$

We continue to argue with exactly the same reasoning as in the proof of (22), replacing  $f$  with  $f'$  and  $F$  with  $F \cap 2^{\ell+5} J$ . This leads to the proof of (31) and thus to

$$\sum_{T \in \mathbf{T}^{\text{main}}} |I_T| \lesssim \mathcal{E}^{-2} |F \cap 2^{\ell+5} J|$$

which is clearly  $\lesssim \mathcal{E}^{-2} 2^\ell |J|$ . Thus (30) is established.

For the complimentary terms we prove better estimates, namely, for  $j = 0, 1, 2, \dots$ ,

$$(32) \quad \sum_{T \in \mathbf{T}_j} |I_T| \lesssim 2^{-\ell-j} \mathcal{E}^{-2} |J|.$$



For each  $T \in \mathbf{T}^j$  we have  $|I| \leq 2^{-j}|I_T| \leq 2^{-j}|J|$  for all  $P \in S(T)$  and thus

$$\sum_{T \in \mathbf{T}^j} |I_T| \lesssim \sum_{T \in \mathbf{T}^j} 2^j \mathcal{E}^{-2} \sum_{P \in S(T): |I| \leq 2^{-j}|J|} |\langle f'', \phi_P \rangle|^2.$$

Since the  $S$  are pairwise disjoint, the right hand side is

$$\lesssim 2^j \mathcal{E}^{-2} \sum_{k \geq j} \sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle f'', \phi_P \rangle|^2.$$

Fixing  $k \geq j$ , we apply Minkowski's inequality to obtain

$$\sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle f'', \phi_P \rangle|^2 \leq \left( \sum_{\substack{K: |K| = 2^{1-k}|J| \\ K \cap 2^{\ell+2}J = \emptyset}} \left( \sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle \mathbb{1}_K f'', \phi_P \rangle|^2 \right)^{1/2} \right)^2$$

where above, we sum over dyadic intervals  $K$  and use the fact that  $f''$  is supported on  $\mathbb{R} \setminus 2^{\ell+5}J$ .

Now note that  $\phi_{P'} = c \exp(2\pi i(c(\omega'_u) - c(\omega_u)) \cdot) \phi_P$  when  $I = I'$ . Thus if  $\mathcal{P}(I_o)$  is a collection of disjoint multitiles with common time interval  $I_o$ , if  $g$  is supported on  $CI_o$  and if  $P_{I_o}$  is any fixed multitile in  $\mathcal{P}(I_o)$  then Bessel's inequality gives

$$\sum_{P \in \mathcal{P}(I_o)} |\langle g, \phi_P \rangle|^2 \lesssim |I_o| \int |g \phi_{P_{I_o}}|^2 dx.$$

We apply this observation to the inner sums in the previous display and obtain the inequality

$$(33) \quad \sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle f'', \phi_P \rangle|^2 \lesssim 2^{-k}|J| \left( \sum_{\substack{K: |K| = 2^{1-k}|J| \\ K \cap 2^{\ell+2}J = \emptyset}} \left( \sum_{\substack{I: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} \|\mathbb{1}_K f'' \phi_{P_I}\|_{L^2}^2 \right)^{1/2} \right)^2$$

where for each  $I$ ,  $P_I$  is any multitile with time interval  $I$ . Since  $|f''| \leq 1$  the bound (14) yields

$$\|\mathbb{1}_K f'' \phi_{P_I}\|_{L^2}^2 \lesssim \left(1 + \frac{\text{dist}(K, I)}{|I|}\right)^{-N}.$$

Applying it with large  $N$  we see that the right hand side of (33) is  $\lesssim 2^{-(k+\ell)(N-4)} 2^{-k}|J|$ . Summing over  $k \geq j$  we obtain inequality (32). This concludes the proof of (18).

*Proof of (20).* For each  $T \in \mathbf{T}$ , let  $S \equiv S(T)$  be the corresponding  $l$ -overlapping tree from the selection algorithm above and recall

$$\sum_{T \in \mathbf{T}} |I_T| (\mathcal{E}/2)^2 \leq \sum_{P \in \bigcup_{T \in \mathbf{T}} S} |\langle f, \phi_P \rangle|^2.$$

Since  $\mathbf{P} = \bigcup_{T' \in \mathbf{T}'} T'$ , the right side above is dominated by

$$(34) \quad \sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \in C_2 \omega_l}} |\langle f, \phi_P \rangle|^2 + \sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \geq \inf C_3 \omega_u}} |\langle f, \phi_P \rangle|^2 \\ + \sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \sup C_2 \omega_l \leq \xi_{T'} < \inf C_3 \omega_u}} |\langle f, \phi_P \rangle|^2.$$

For each  $T' \in \mathbf{T}'$  the set of tiles  $P \in T' \cap \bigcup_{T \in \mathbf{T}} S(T)$  with the property that  $\xi_{T'} \in C_2\omega_l$  is by definition an  $l$ -overlapping tree. Thus, since  $\mathbf{P}$  has energy bounded by  $\mathcal{E}$ , we estimate the first term in (34)

$$\sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \in C_2\omega_l}} |\langle f, w_u \rangle|^2 \leq \sum_{T' \in \mathbf{T}'} \mathcal{E}^2 |I_{T'}|.$$

For the estimation of the second term in (34) we observe that any fixed multitile forms an  $l$ -overlapping tree (with respect to some top data), and we use the energy bound for fixed tiles. We observe that the rectangles  $\{I \times [\inf C_3\omega_u, \sup C_2\omega_u) : P \in \bigcup_{T \in \mathbf{T}} S\}$  are pairwise disjoint. Indeed if  $P \in S$ ,  $P' \in S$  this follows since  $S$  is  $l$ -overlapping (cf. Observation 4.1). If  $P \in S_{\kappa_1}$  and  $P' \in S_{\kappa_2}$   $\kappa_1 < \kappa_2$  then  $\xi_{S_{\kappa_1}} \geq \xi_{S_{\kappa_2}}$  and an overlap of  $[\inf C_3\omega_u, \sup C_2\omega_u)$  and  $[\inf C_3\omega'_u, \sup C_2\omega'_u)$  would imply that  $P'$  belongs to the maximal tree with the same top data as  $S_{\kappa_1}$ , i.e. this would imply that  $P' \in T_{\kappa_1}$  which is disjoint from  $S_{\kappa_2}$ .

This allows us to estimate for any fixed  $T' \in \mathbf{T}'$

$$\sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \geq \inf C_3\omega_u}} |\langle f, \phi_P \rangle|^2 \leq \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \geq \inf C_3\omega_u}} \mathcal{E}^2 |I| \leq \mathcal{E}^2 |I_{T'}|.$$

Now sum over  $T' \in \mathbf{T}'$  and it follows that the middle term in (34) is  $\lesssim \sum_{T' \in \mathbf{T}'} \mathcal{E}^2 |I_{T'}|$  and summing

For the estimation of the third term in (34) we begin with a preliminary observation, also related to the selection of the  $S$ . Suppose  $P \in T' \cap S$ ,  $\tilde{P} \in T' \cap \tilde{S}$  where  $T, \tilde{T} \in \mathbf{T}$  and

$$\xi_{T'} \in [\sup C_2\omega_l, \inf C_3\omega_u) \cap [\sup C_2\tilde{\omega}_l, \inf C_3\tilde{\omega}_u),$$

and suppose  $I \subset \tilde{I}$  and  $P \neq \tilde{P}$ . From (16) we have  $I \subsetneq \tilde{I}$ . We also have  $\inf C_2\tilde{\omega}_l < \sup C_2\omega_l$  since otherwise it would follow that  $\tilde{S}$  was selected prior to  $S$  and hence  $P \in \tilde{T}$  which is impossible. From (15), we have  $\inf C_2\tilde{\omega}_l \geq \sup C_3\omega_l$  and so  $P$  is in the maximal  $l$ -overlapping tree contained in  $T'$  with top data  $(\tilde{I}, \inf C_2\tilde{\omega}_l)$ .

For each  $T' \in \mathbf{T}'$  let  $T''$  be the collection of multitiles  $P \in T' \cap \bigcup_{T \in \mathbf{T}} S$  with  $\xi_{T'} \in [\sup C_2\omega_l, \inf C_3\omega_u)$  and  $I$  maximal among such multitiles. Then

$$\sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \sup C_2\omega_l \leq \xi_{T'} < \inf C_3\omega_u}} |\langle f, \phi_P \rangle|^2 \leq \sum_{T' \in \mathbf{T}'} \sum_{P'' \in T''} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \sup C_2\omega_l \leq \xi_{T'} < \inf C_3\omega_u \\ I \subset I''}} |\langle f, \phi_P \rangle|^2$$

Considering the discussion in the preceding paragraph, we may apply the energy bound to the maximal  $l$ -overlapping tree contained in  $T'$  with top data  $(\tilde{I}, \inf C_2\tilde{\omega}_l)$ . Since the  $I''$  are disjoint subintervals of  $I_{T'}$  we see that the right side in the last display is bounded by

$$\sum_{T' \in \mathbf{T}'} \sum_{P'' \in T''} 2\mathcal{E}^2 |I''| \leq \sum_{T' \in \mathbf{T}'} 2\mathcal{E}^2 |I_{T'}|.$$

This completes the proof of (20).  $\square$

The proposition below is for use in tandem with Proposition 4.3.

**Proposition 4.4.** *Let  $\mathbf{P}$  be a collection of multitiles and  $\mu > 0$ . Then, there is a collection of trees  $\mathbf{T}$  such that*

$$(35) \quad \sum_{T \in \mathbf{T}} |I_T| \lesssim \mu^{-r'} |E|$$

and such that

$$\text{density}(\mathbf{P} \setminus \cup_{T \in \mathbf{T}} T) \leq \mu/2.$$

*Proof.* We select trees through an iterative procedure. Suppose that trees  $T_j, T_j^+, T_j^-$  have been chosen for  $j = 1, \dots, k$ . Let

$$\mathbf{P}_k = \mathbf{P} \setminus \bigcup_{j=1}^k T_j \cup T_j^+ \cup T_j^-.$$

If  $\text{density}(\mathbf{P}_k) \leq \mu/2$  then we terminate the procedure and set

$$\mathbf{T} = \{T_1, T_1^+, T_1^-, \dots, T_k, T_k^+, T_k^-\}.$$

Otherwise, we may find a nonempty tree  $T \subset \mathbf{P}_k$  such that

$$(36) \quad \frac{1}{|I_T|} \int_E \left(1 + \frac{|x - c(I_T)|}{|I_T|}\right)^{-4} \sum_{k: \xi_{k-1}(x) \in \omega_T} |a_k(x)|^{r'} dx > (\mu/2)^{r'}.$$

Choose  $T_{k+1} \subset \mathbf{P}_k$  so that  $|I_{T_{k+1}}|$  is maximal among all nonempty trees contained in  $\mathbf{P}_k$  which satisfy (36), and so that  $T_{k+1}$  is the maximal, with respect to inclusion, tree contained in  $\mathbf{P}_k$  with top data  $(I_{T_{k+1}}, \xi_{T_{k+1}})$ . Let  $T_{k+1}^+ \subset \mathbf{P}_k$  be the maximal tree contained in  $\mathbf{P}_k$  with top data  $(I_{T_{k+1}}, \xi_{T_{k+1}} + (C_2 - 1)/(2|I_{T_{k+1}}|))$  and  $T_{k+1}^- \subset \mathbf{P}_k$  be the maximal tree contained in  $\mathbf{P}_k$  with top data  $(I_{T_{k+1}}, \xi_{T_{k+1}} - (C_2 - 1)/(2|I_{T_{k+1}}|))$ . Since each  $T_j$  is nonempty and  $\mathbf{P}$  is finite, this process will eventually stop.

To prove (35), it will suffice to verify

$$(37) \quad \sum_j |I_{T_j}| \lesssim \mu^{-r'} |E|.$$

To this end, we first observe that the tiles  $I_{T_j} \times \omega_{T_j}$  are pairwise disjoint. Indeed, suppose that  $(I_{T_j} \times \omega_{T_j}) \cap (I_{T_{j'}} \times \omega_{T_{j'}}) \neq \emptyset$  and  $j < j'$ . Then, by the first maximality condition, we have  $|I_{T_j}| \geq |I_{T_{j'}}|$  and so  $I_{T_{j'}} \subset I_{T_j}$  and  $|\omega_{T_j}| \leq |\omega_{T_{j'}}|$ . From the latter inequality, it follows that for every  $P \in T_{j'}$ , either  $\omega_{T_j} \subset \omega_m$ ,  $\omega_{T_j^+} \subset \omega_m$ , or  $\omega_{T_j^-} \subset \omega_m$ . Thus,  $T_{j'} \subset T_j \cup T_j^+ \cup T_j^-$  which contradicts the selection algorithm.

Breaking the integral up into pieces and applying a pigeonhole argument, it follows from (36) that for each  $j$  there is a positive integer  $\ell_j$  such that

$$(38) \quad |I_{T_j}| \leq C 2^{-3\ell_j} \mu^{-r'} \int_{E \cap 2^{\ell_j} I_{T_j}} \sum_{k: \xi_{k-1}(x) \in \omega_{T_j}} |a_k(x)|^{r'} dx.$$

For each  $\ell$  we let  $\mathbf{T}^{(\ell)} = \{T_j : \ell_j = \ell\}$  and choose elements of  $\mathbf{T}^{(\ell)}$ :  $T_1^{(\ell)}, T_2^{(\ell)}, \dots$  and subsets of  $\mathbf{T}^{(\ell)}$ :  $\mathbf{T}_1^{(\ell)}, \mathbf{T}_2^{(\ell)}, \dots$  as follows. Suppose  $T_j^{(\ell)}$  and  $\mathbf{T}_j^{(\ell)}$  have been chosen

for  $j = 1, \dots, k$ . If  $\mathbf{T}^{(\ell)} \setminus \bigcup_{j=1}^k \mathbf{T}_j^{(\ell)}$  is empty, then terminate the selection procedure. Otherwise, let  $T_{k+1}^{(\ell)}$  be an element of  $\mathbf{T}^{(\ell)} \setminus \bigcup_{j=1}^k \mathbf{T}_j^{(\ell)}$  with  $|I_{T_{k+1}^{(\ell)}}|$  maximal, and let

$$\mathbf{T}_{k+1}^{(\ell)} = \{T \in \mathbf{T}^{(\ell)} \setminus \bigcup_{j=1}^k \mathbf{T}_j^{(\ell)} : (2^\ell I_T \times \omega_T) \cap (2^\ell I_{T_{k+1}^{(\ell)}} \times \omega_{T_{k+1}^{(\ell)}}) \neq \emptyset\}.$$

By construction,  $\mathbf{T}^{(\ell)} = \bigcup_j \mathbf{T}_j^{(\ell)}$  and so

$$(39) \quad \sum_{T \in \mathbf{T}^{(\ell)}} |I_T| \leq \sum_j \sum_{T \in \mathbf{T}_j^{(\ell)}} |I_T|.$$

Using the fact that the tiles  $I_{T_j} \times \omega_{T_j}$  are pairwise disjoint, and (twice) the fact that  $|I_T| \leq |I_{T_j^{(\ell)}}|$  for every  $T \in \mathbf{T}_j^{(\ell)}$ , we see that for each  $j$

$$\sum_{T \in \mathbf{T}_j^{(\ell)}} |I_T| \lesssim 2^\ell |I_{T_j^{(\ell)}}|.$$

From (38), we thus see that the right side of (39) is dominated by a constant times

$$\begin{aligned} & 2^{-2\ell} \mu^{-r'} \int_E \sum_j \mathbb{1}_{2^\ell I_{T_j^{(\ell)}}}(x) \sum_{k: \xi_{k-1}(x) \in \omega_{T_j^{(\ell)}}} |a_k(x)|^{r'} dx \\ &= 2^{-2\ell} \mu^{-r'} \int_E \sum_k |a_k(x)|^{r'} \# \{j : (x, \xi_{k-1}(x)) \in 2^\ell I_{T_j^{(\ell)}} \times \omega_{T_j^{(\ell)}}\} dx \\ &\leq 2^{-2\ell} \mu^{-r'} |E|. \end{aligned}$$

where we used the disjointness of the rectangles  $2^\ell I_{T_j^{(\ell)}} \times \omega_{T_j^{(\ell)}}$ , and  $\sum_{k=1}^K |a_k(x)|^{r'} \leq 1$ . Summing over  $\ell$ , we obtain (37).  $\square$

## 5. THE TREE ESTIMATE

In this section we prove the basic estimate for the model operators in the special case where the collection of multitiles is a tree. In what follows we use the notation  $\mathcal{V}^r A f(x)$  for the  $r$ -variation of  $k \mapsto A_k f(x)$ , for a given family of operators  $A_k$  indexed by  $k \in \mathbb{N}$ .

An essential tool introduced to harmonic analysis by Bourgain [1] is *Lépingle's inequality* for martingales ([22]). Consider the martingale of dyadic averages

$$\mathbb{E}_k[f](x) = \frac{1}{|I_k(x)|} \int_{I_k(x)} f(y) dy$$

where  $I_k(x)$  is the dyadic interval of length  $2^k$  containing  $x$ . It is a special case of Lépingle's inequality that

$$(40) \quad \|\mathcal{V}^r \mathbb{E}_{(\cdot)}[f]\|_{L^p} \leq C_{p,r} \|f\|_{L^p}$$

whenever  $1 < p < \infty$  and  $r > 2$ . Simple proofs (based on jump inequalities) have been obtained in [1] and [27] (see also [9], [15] for other expositions). Inequality (40) has been extended to various families of convolution operators ([1], [14], [2], [15]). Let

$\psi$  be a Schwartz function on  $\mathbb{R}$  with  $\int \psi = 1$ , for each  $k$  let  $\psi_k = 2^{-k}\psi(2^{-k}\cdot)$  and let  $A_k f(x) = \psi_k * f$ . Then

$$(41) \quad \|\mathcal{V}^r A f\|_{L^p} \leq C'_{p,r} \|f\|_{L^p}, \quad 1 < p < \infty, \quad r > 2,$$

follows from (40); the essential tool is the square function estimate

$$(42) \quad \left\| \left( \sum_{k=-\infty}^{\infty} |A_k f - \mathbb{E}_k[f]|^2 \right)^{1/2} \right\|_p, \quad 1 < p < \infty.$$

For a proof we refer to [14] or [15].

We shall use (41) to prove the following estimate in terms of energy and density. The bound will be applied in Section 7 with  $q = r'$  and  $q = 1$ .

**Proposition 5.1.** *Let  $\mathcal{E} > 0$ ,  $0 < \mu < 3$  and let  $T$  be a tree with energy bounded above by  $\mathcal{E}$  and density bounded above by  $\mu$ . Then, for each  $1 \leq q \leq 2$*

$$(43) \quad \left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}.$$

Furthermore, for  $\ell \geq 0$  we have

$$(44) \quad \left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(\mathbb{R} \setminus 2^\ell I_T)} \lesssim 2^{-\ell(N-10)} \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}.$$

We remark that the bounds above also hold for  $2 < q < \infty$ , but the result for this range of exponents is not needed for our purposes; it requires an additional  $L^p$  estimate for  $\sum_{P \in T} \langle f, \phi_P \rangle \phi_P$ .

*Proof.* We begin with some preliminary reduction. Every tree can be split into an  $l$ -lacunary tree and an  $l$ -overlapping tree and it suffices to prove the asserted estimate for these cases.

If the tree is  $l$ -overlapping we introduce further decompositions. By breaking up  $T$  into a bounded number of subtrees we may and shall assume without loss of generality that for each  $P \in T$ ,

$$(45) \quad \xi_T \in \omega_l + j|\omega_l| \text{ for some integer } j \text{ with } |j| \leq C_2.$$

Moreover we shall assume, for every  $l$ -overlapping tree  $T$ , that either  $\xi_T \leq \inf \omega_l$  for every  $P \in T$  (in which case we refer to  $T$  as  $l^-$ -overlapping) or  $\xi_T > \inf \omega_l$  for every  $P \in T$  (in which case we refer to  $T$  as  $l^+$ -overlapping). Every  $l$ -overlapping tree can be split into an  $l^-$ -overlapping and an  $l^+$ -overlapping tree. For the remainder of the proof, we assume without loss of generality that  $T$  is either  $l$ -lacunary or  $l^+$ -overlapping or  $l^-$ -overlapping, and that for the last two categories property (45) is satisfied.

Let  $\mathbf{J}$  be the collection of dyadic intervals  $J$  which are maximal with respect to the property that  $I \not\subset 3J$  for every  $P \in T$ .

Our first goal is to prove that for each  $J \in \mathbf{J}$

$$(46) \quad \left\| \sum_{P \in T: |I| < C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |J|^{1/q} \left( 1 + \frac{\text{dist}(I_T, J)}{|I_T|} \right)^{-(N-6)}$$

where  $C'' \geq 1$  is a constant to be determined later; we shall see that  $C'' = 8C_1(C_2 - 1)^{-1}$  is an admissible choice.

By Hölder's inequality, we may assume that  $q \geq r'$ . Fix  $P \in T$  with  $|I| \leq C''|J|$ . From the energy bound, we have

$$(47) \quad \|\langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E\|_{L^q(J)} \lesssim \mathcal{E} \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^{-N} \|a_P \mathbb{1}_E\|_{L^q(J)}.$$

From the density bound applied to  $\approx 1/(C_2 - 1)$  nonempty trees, each with top time interval  $I$ , we obtain

$$\frac{1}{|I|} \int_E \left(1 + \frac{|x - c(I)|}{|I|}\right)^{-4} \sum_{k: \xi_{k-1}(x) \in \omega_l} |a_k(x)|^{r'} dx \lesssim \mu^{r'}.$$

Since  $I \not\subset 3J$ , it follows that  $1 + |x - y|/|I| \leq C(1 + \text{dist}(I, J)/|I|)$  for every  $x \in J$  and  $y \in I$ . Thus

$$\|a_P \mathbb{1}_E\|_{L^q(J)}^q \leq \|a_P \mathbb{1}_E\|_{L^{r'}(J)}^{r'} \lesssim \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^4 |I| \mu^{r'}$$

where, above, we use the fact that  $|a_P| \leq 1$ . Thus we can replace (47) by

$$\|\langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E\|_{L^q(J)} \lesssim \mathcal{E} \mu^{r'/q} |I|^{1/q} \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^{-(N-4)}.$$

Summing this estimate and using the fact that  $T$  is a tree, we have

$$\left\| \sum_{P \in T: |I|=2^{-k}|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)} \lesssim \mathcal{E} \mu^{r'/q} (2^{-k}|J|)^{1/q} \left(1 + \frac{\text{dist}(I_T, J)}{|I_T|}\right)^{-(N-6)}$$

and summing over  $k$  gives (46).

We now use (46) to prove (44) for  $\ell \geq 4$ . Indeed, using the maximality of each  $J$ , we see that if  $\ell \geq 4$  and  $J \cap (\mathbb{R} \setminus 2^\ell I_T) \neq \emptyset$  then  $\text{dist}(I_T, J) \geq |J|/2$  and  $|J| \geq 2^{\ell-3}|I_T|$ . It thus follows from (46) that

$$\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)} \lesssim (|I_T|/|J|) (\text{dist}(I_T, J)/|J|)^{-2} \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q} 2^{-\ell(N-10)}$$

whenever  $J \cap (\mathbb{R} \setminus 2^\ell I_T) \neq \emptyset$ . Summing over all  $J$ , we thus obtain (44) for  $\ell \geq 4$ .

It remains to prove

$$\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(16I_T)} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}$$

which, by (46), follows from

$$(48) \quad \left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(16I_T)} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}.$$

Again, by Hölder's inequality, we may assume that  $q \geq r'$ . Let

$$\Omega_J = \bigcup_{P \in T: |I| \geq C''|J|} \omega_l.$$

The first step in the proof of (48) will be to demonstrate

$$(49) \quad \int_{J \cap E} \sum_{k: \xi_{k-1}(x) \in \Omega_J} |a_k(x)|^{r'} dx \lesssim \mu^{r'} |J|.$$

By the maximality of  $J$  there is a multitile  $P^* = (I^*, \omega_u^*) \in T$  with  $I^* \subset 3\tilde{J}$  where  $\tilde{J}$  is the dyadic double of  $J$ . This implies that there is a dyadic interval  $J'$  with  $|J| \leq |J'| \leq 4|J|$

and  $\text{dist}(J, J') \leq |J|$  and  $I^* \subset J'$ . We wish to apply the density bound assumption to a tree consisting of the one multitile  $P^*$ , with top interval  $J'$  and suitable choice of the top frequency. We distinguish the cases that  $T$  is  $l^+$ -overlapping,  $l^-$ -overlapping, or  $l$ -lacunary.

If  $T$  is  $l^+$ -overlapping then  $T' = (\{P^*\}, \xi_T, J')$  is a tree. For every  $P \in T$  we have  $\omega_l \subset [\xi_T - \frac{C_1}{2|I|}, \xi_T + \frac{C_1}{2|I|}]$  and thus, with  $|I| \geq C''|J|$ ,  $\omega_l \subset [\xi_T - \frac{C_1}{2C''|J|}, \xi_T + \frac{C_1}{2C''|J|}]$ . Thus with the choice of  $C'' \geq \frac{8C_1}{C_2-1}$  we have  $\Omega_J \subset \omega_{T'}$ .

If  $T$  is  $l^-$ -overlapping then  $T' = (\{P^*\}, \xi_T + \frac{C_2-1}{4|J|}, J')$  is a tree. Using that  $T$  is  $l^-$ -overlapping, we see that  $\omega_l \subset [\xi_T, \xi_T + \frac{C_2}{2C''|J|}]$  for every  $P \in T$  with  $|I| \geq C''|J|$ . Thus, by choosing  $C'' \geq \frac{8C_2}{C_2-1}$ , we have  $\Omega_J \subset \omega_{T'}$ .

If  $T$  is  $l$ -lacunary then  $T' = (\{P^*\}, \xi_T - \frac{C_2-1}{4|J|}, J')$  is a tree. Using that  $T$  is  $l$ -lacunary, we see that  $\omega_l \subset [\xi_T - \frac{C_1}{2C''|J|}, \xi_T]$  for every  $P \in T$  with  $|I| \geq C''|J|$ . Thus, by choosing  $C'' \geq 8\frac{C_1}{C_2-1}$  as in the first case we have  $\Omega_J \subset \omega_{T'}$ .

In any of the three cases, the density bound gives

$$\frac{1}{|J'|} \int_E \left(1 + \frac{|x - c(J')|}{|J'|}\right)^{-4} \sum_{k: \xi_{k-1}(x) \in \omega_{T'}} |a_k(x)|^{r'} dx \leq \mu^{r'}$$

and hence (49).

We now show that if  $T$  is  $l$ -lacunary then (48) follows from (49). We start by observing that for each  $x$  there is at most one integer  $m$  and at most one integer  $k$  such that there exists a  $P \in T$  with  $|I| = 2^m$ ,  $\xi_{k-1}(x) \in \omega_l$ , and  $\xi_k(x) \in \omega_h$ . Indeed, suppose such a  $P$  exists, and  $P' \in T$  with  $|I'| \geq |I|$ . If  $|I| = |I'|$  the uniqueness of  $k$  is obvious. Suppose  $|I'| > |I|$ . Since  $T$  is  $l$ -lacunary, we have  $\inf(\omega_{l'}) > \sup(\omega_l)$  by (15), and so  $\xi_{k-1}(x) < \inf(\omega_{l'})$ . We also have  $\xi_k(x) \geq \inf(\omega_h) > \sup(\omega_T) > \sup(\omega_{l'})$  since  $C_2\omega_u \cap \omega_h = \emptyset$  and  $\xi_T \notin C_2\omega_{l'}$ . It follows that there is no  $k'$  with  $\xi_{k'-1}(x) \in \omega_{l'}$ .

Now let  $a(x) = a_k(x)$  if there exists an  $m(x)$  as in the previous paragraph with  $2^{m(x)} \geq C''|J|$ , and  $a(x) = 0$  otherwise. We then have

$$\left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)}^q \leq \int_{J \cap E} \left( |a(x)| \sum_{P \in T: |I|=2^{m(x)}} |\langle f, \phi_P \rangle \phi_P(x)| \right)^q dx.$$

From the energy bound (applied to multitiles) and the bound (14) for  $|\phi_P|$ , the right side is bounded by

$$\int_{J \cap E} \left[ |a(x)| \sum_{\substack{P \in T: \\ |I|=2^{m(x)}}} \mathcal{E} |I| w_I(x) \right]^q dx \lesssim \int_{J \cap E} \left[ |a(x)| \sum_{\substack{P \in T: \\ |I|=2^{m(x)}}} \mathcal{E} \left(1 + \frac{|x - c(I)|}{|I|}\right)^{-N} \right]^q dx.$$

Noting that  $\sum_{P \in T: |I|=2^{m(x)}} (1 + |x - c(I)|/|I|)^{-N} \leq C$ , we see that the last display is

$$\lesssim \mathcal{E}^q \int_{J \cap E} |a(x)|^q dx \lesssim \mathcal{E}^q \int_{J \cap E} \left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} |a_k(x)|^{r'} \right)^{q/r'} dx,$$



by our choice of  $a(x)$ . Using (49),  $q \geq r'$ , and the fact that  $\sum |a_k(x)|^{r'} \leq 1$ , the right hand side is  $\lesssim \mathcal{E}^q \mu^{r'} |J|$ . We may now sum over those  $J \in \mathbf{J}$  which satisfy  $J \cap 16I_T \neq \emptyset$  and (48) follows for  $l$ -lacunary trees.

It remains to prove (48) for the case when  $T$  is  $l$ -overlapping and satisfies condition (45). For each  $J \in \mathbf{J}$ , and each  $x \in J \cap E$  we have by Hölder's inequality

$$(50) \quad \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P(x) a_P(x) = \sum_k a_k(x) \sum_{\substack{P \in T: |I| \geq C''|J| \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) \\ \leq \left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} |a_k(x)|^{r'} \right)^{1/r'} \left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} \left| \sum_{\substack{P \in T: |I| \geq C''|J| \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) \right|^r \right)^{1/r}.$$

Now let  $\psi$  be a Schwartz function with  $\widehat{\psi}(\xi) = 1$  for  $|\xi| \leq C_1 + C_3$  and  $\widehat{\psi}(\xi) = 0$  for  $|\xi| \geq 2C_1$ . Define  $\psi_\ell = 2^{-\ell} \psi(2^{-\ell} \cdot)$ , and set

$$e_T(x) = e^{2\pi i \xi_T x}.$$

We will show that for any  $x$  and any  $k$  with  $\xi_{k-1}(x) < \xi_k(x)$ , there exist integers  $\ell_1, \ell_2$  depending on  $x$  and  $k$ , such that  $2^{\ell_1} \geq |J|$  and

$$(51) \quad \sum_{\substack{P \in T: |I| \geq C''|J|, \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) = (e_T(\psi_{\ell_1} - \psi_{\ell_2})) * \left[ \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right](x).$$

From (15) we have, for each  $\ell$  such that  $2^\ell = |I|$  for some multitile  $P$ ,

$$(e_T \psi_\ell) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(x) = \sum_{P \in T: |I| \geq 2^\ell} \langle f, \phi_P \rangle \phi_P(x).$$

Thus, to prove (51) it will suffice to show that there exist integers  $\ell_1$  and  $\ell_2$  such that

$$(52) \quad \{P \in T : |I| \geq C''|J|, \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h\} = \{P \in T : 2^{\ell_1} \leq |I| \leq 2^{\ell_2}\}.$$

Again using (15), we see that for  $P, P' \in T$  with  $|I| < |I'|$  we have  $\inf \omega'_h < \inf \omega_h$ , and if we are in the setting of  $\rho$ -multitiles where  $\rho$  is a 1-index, we have the stronger inequality  $\sup \omega'_h < \inf \omega_h$ . Thus, (52) will follow after finding  $\ell_1$  and  $\ell_2$  with

$$(53) \quad \{P \in T : \xi_{k-1}(x) \in \omega_l\} = \{P \in T : 2^{\ell_1} \leq |I| \leq 2^{\ell_2}\}.$$

Here we use assumption (45). The displayed equation follows when  $|j| > 1$  from the fact that  $\omega_l \cap \omega'_l = \emptyset$  if  $P, P' \in T$  and  $|I| < |I'|$ ; it follows when  $j = 0$  from the fact that the intervals  $\{\omega_l : P \in T\}$  are nested. Finally, when  $j = \pm 1$  it follows from the property that if  $P, P', P'' \in T$ ,  $|I| \leq |I'| \leq |I''|$  and  $\omega_l \cap \omega''_l \neq \emptyset$  then  $\omega'_l \subset \omega'_l \subset \omega_l$ . Thus we have established (53) and consequently (52) and (51).

Using (51), we have

$$\left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} \left| \sum_{\substack{P \in T: |I| \geq C''|J|, \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) \right|^r \right)^{1/r} \\ \leq \| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(x) \|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))}$$

where the notation refers to the variation norm with respect to the variable  $k$ , restricted to  $\{k \in \mathbb{Z} : |k| \geq \log_2(|J|)\}$ . For  $\log_2(|J|) \leq k_1 < k_2$ , we have

$$(e_T(\psi_{k_1} - \psi_{k_2})) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P = (e_T \psi_{C+\log_2(|J|)}) * (e_T(\psi_{k_1} - \psi_{k_2})) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P$$

and so, for  $x \in J$

$$\begin{aligned} & \left\| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(x) \right\|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))} \\ & \lesssim \sup_{x \in J} \sup_{R \geq |J|} \frac{2}{|R|} \int_{x-R}^{x+R} \left\| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(y) \right\|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))} dy. \end{aligned}$$

We now integrate the  $q$ th power of the expressions in (50) over  $E \cap J$  and obtain

$$\begin{aligned} & \left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)}^q \leq \int_{J \cap E} \left( \sum_{k: \xi_{k-1}(t) \in \Omega_J} |a_k(t)|^{r'} dt \right)^{q/r'} \times \\ & \sup_{x \in J} \sup_{R \geq |J|} \frac{2}{|R|} \int_{x-R}^{x+R} \left\| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(y) \right\|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))} dy. \\ & \leq \mu^{r'} \int_J \left| \mathcal{M} \left[ \left\| \psi_k * (e_T^{-1} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P) \right\|_{V_k^r} \right](x) \right|^q dx \end{aligned}$$

where  $\mathcal{M}$  is the Hardy-Littlewood maximal operator. For the last inequality we have used  $q \geq r'$ ,  $\sum |a_k(x)|^{r'} \leq 1$  and inequality (49). Summing over  $J \in \mathbf{J}$  gives

$$\begin{aligned} & \left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(16I_T)}^q \\ & \lesssim \mu^{r'} \left\| \mathcal{M} \left[ \left\| \psi_k * (e_T^{-1} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P) \right\|_{V_k^r} \right](x) \right\|_{L_x^q(16I_T)}^q. \end{aligned}$$

Since  $q \leq 2$ , it follows from Hölder's inequality that the right side above is

$$\lesssim \mu^{r'} |I_T|^{(2-q)/2} \left\| \mathcal{M} \left[ \left\| \psi_k * (e_T^{-1} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P) \right\|_{V_k^r} \right](x) \right\|_{L_x^2(16I_T)}^q.$$

Applying the variation estimate (41) with  $p = 2$  and the  $L^2$  estimate for  $\mathcal{M}$  one sees that the display above is

$$\lesssim \mu^{r'} |I_T|^{(2-q)/2} \left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right\|_{L^2}^q.$$

To finish the proof, it only remains to see that  $\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right\|_{L^2}^2 \lesssim \mathcal{E}^2 |I_T|$ . The left side of this inequality is dominated by

$$\sum_{P \in T} \sum_{P' \in T} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \leq 2 \sum_{P \in T} |\langle f, \phi_P \rangle|^2 \sum_{P' \in T} |\langle \phi_P, \phi_{P'} \rangle|.$$

Since  $T$  is an  $l$ -overlapping tree, we have  $\langle \phi_P, \phi_{P'} \rangle$  unless  $|I| = |I'|$ , in which case, we have  $|\langle \phi_P, \phi_{P'} \rangle| \lesssim (1 + \text{dist}(I, I')/|I|)^{-N}$ . Therefore we obtain the estimate

$$\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right\|_{L^2}^2 \lesssim \sum_{P \in T} |\langle f, \phi_P \rangle|^2 \leq C \mathcal{E}^2 |I_T|.$$

This concludes the proof of (48) and thus the proof of the proposition.  $\square$

## 6. TWO AUXILIARY ESTIMATES

Before we give the argument on how to decompose our operators into trees with suitable energy and density bounds we need two auxiliary estimates.

The following proposition can be found in [33], p. 12, or as a special case of a lemma from [12].

**Proposition 6.1.** *Let  $T$  be an  $l$ -overlapping tree. Let  $\lambda > 0$  and  $\Omega_{\lambda,D} = \{\mathcal{M}_D[\mathbb{1}_F] > \lambda\}$  where  $\mathcal{M}_D$  is the maximal dyadic average operator. Then*

$$\frac{1}{|I_T|} \sum_{P \in T: I \not\subset \Omega_{\lambda,D}} |\langle f, \phi_P \rangle|^2 \lesssim \lambda^2.$$

The second auxiliary estimate is the special case of an estimate from [12], but we will provide a proof for convenience.

**Proposition 6.2.** *Let  $\mathbf{P}$  be a finite set of multitiles, and let  $\lambda > 0$ ,  $F \subset \mathbb{R}$ , and  $|f| \leq \mathbb{1}_F$ . Let  $\Omega_\lambda = \{\mathcal{M}[\mathbb{1}_F] > \lambda\}$ . Then*

$$(54) \quad \left\| \sum_{P \in \mathbf{P}: I \subset \Omega} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus \Omega)} \lesssim \frac{|F|}{\lambda^{1/r}}.$$

*Proof.* Fix  $\ell$  and let  $I_\ell \subset \Omega_\lambda$  be a dyadic interval satisfying

$$(55) \quad 2^\ell I_\ell \subset \Omega_\lambda \text{ and } 2^{\ell+1} I_\ell \not\subset \Omega_\lambda.$$

By Minkowski's inequality, we estimate

$$\begin{aligned} & \left( \sum_{P: I = I_\ell} |\langle f, \phi_P \rangle|^2 \right)^{1/2} \\ & \leq \left( \sum_{P: I = I_\ell} |\langle \mathbb{1}_{4I_\ell} f, \phi_P \rangle|^2 \right)^{1/2} + \sum_{j=2}^{\infty} \left( \sum_{P: I = I_\ell} |\langle \mathbb{1}_{2^{j+1}I_\ell \setminus 2^j I_\ell} f, \phi_P \rangle|^2 \right)^{1/2}. \end{aligned}$$

Using orthogonality, the right hand side is bounded by

$$(4|I_\ell|)^{1/2} \|\mathbb{1}_{4I_\ell} f \phi_{P_0}\|_{L^2} + \sum_{j=2}^{\infty} (2^{j+1}|I_\ell|)^{1/2} \|\mathbb{1}_{2^{j+1}I_\ell \setminus 2^j I_\ell} f \phi_{P_0}\|_{L^2}$$

where  $P_0$  is any multtile with  $I = I_\ell$ . Applying the bounds (14) and  $|f| \leq \mathbb{1}_F$ , we see that the last display is

$$\lesssim |F \cap 4I_\ell|^{1/2} + \sum_{j=2}^{\infty} C 2^{-j(N-1)} |F \cap 2^{j+1} I_\ell|^{1/2}.$$

Since  $2^{\ell+1} I_\ell \not\subset \Omega_\lambda$ , we have  $|F \cap 2^{j+1} I_\ell| \leq C 2^{\max(\ell, j)} |I_\ell| \lambda$  for each  $j$ . Thus, the last display is  $\lesssim (2^\ell \lambda |I_\ell|)^{1/2}$  and we have proved

$$\left( \sum_{P: I = I_\ell} |\langle f, \phi_P \rangle|^2 \right)^{1/2} \lesssim (2^\ell \lambda |I_\ell|)^{1/2}.$$

Similarly,

$$\sup_{P: I = I_\ell} |\langle f, \phi_P \rangle| \lesssim 2^\ell \lambda |I_\ell|^{1/2}$$

and so, by interpolation,

$$(56) \quad \left( \sum_{P:I=I_\ell} |\langle f, \phi_P \rangle|^r \right)^{1/r} \lesssim (2^\ell \lambda)^{1/r'} |I_\ell|^{1/2}$$

whenever  $2 \leq r \leq \infty$ . For each  $\xi, I_\ell$  there is at most one  $P \in \mathbf{P}$  with  $\xi \in \omega_I$  and  $I = I_\ell$ . Thus, using the fact that, for each  $x$ ,  $\sum_{k=1}^K |a_k(x)|^{r'} \leq 1$ , we see that

$$\left\| \sum_{P \in \mathbf{P}: I=I_\ell} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus \Omega_\lambda)} \lesssim (2^\ell \lambda)^{1/r'} |I_\ell|^{1/2} \|\phi_{P_0}\|_{L^1(\mathbb{R} \setminus \Omega_\lambda)}$$

where  $P_0$  is any multitile with  $I_0 = I_\ell$ . Using the fact that  $2^\ell I_\ell \subset \Omega_\lambda$ , it follows that the right side above is

$$\lesssim 2^{-\ell(N-2)} \lambda^{1/r'} |I_\ell|.$$

For  $\ell \geq 0$  let  $\mathcal{I}_\ell$  be the set of all dyadic intervals satisfying (55). If  $I \in \mathcal{I}_\ell$  then for each  $j > 0$  there are at most 2 intervals  $I' \in \mathcal{I}_\ell$  with  $I' \subset I$  and  $|I'| = 2^{-j}|I|$ . By considering the collection of maximal dyadic intervals in  $\mathcal{I}_\ell$ , one sees that

$$\sum_{I \in \mathcal{I}_\ell} |I| \lesssim |\Omega_\lambda|$$

Thus,

$$\left\| \sum_{P \in \mathbf{P}: I \in \mathcal{I}_\ell} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus \Omega_\lambda)} \lesssim 2^{-\ell(N-2)} \lambda^{1/r'} |\Omega_\lambda|.$$

Summing over  $\ell$  and applying the weak-type 1-1 estimate for  $\mathcal{M}$  then gives (54).  $\square$

## 7. CONCLUSION OF THE PROOF

Let  $r > 2$  and  $r' \leq p < \frac{2r}{r-2}$ . We shall conclude the proof of (13), with  $1 \leq |E| \leq 2$ , and thus of Theorem 1.2. It will then suffice, by Chebyshev's inequality, to show

$$(57) \quad \int_{E \setminus G} |S^\rho[f](x)| dx \leq C |F|^{1/p}$$

for any measurable set  $E$  with  $1 \leq |E| \leq 2$ ,  $|f| \leq \mathbb{1}_F$  and some exceptional set  $G = G(E, F)$  with  $|G| \leq 1/4$ .

We shall repeatedly apply Propositions 4.3 and 4.4. By Lemma 4.2 the density of  $\mathbf{P}$  (with respect to the set  $E$  above) and the energy (with respect to  $f$ ) are bounded by a universal constant  $C$ .

We distinguish between the cases  $|F| > 1$  and  $|F| \leq 1$  and first consider the case when  $|F| > 1$ . Repeatedly applying Propositions 4.3 and 4.4 we write  $\mathbf{P}$  as the disjoint union

$$\mathbf{P} = \bigcup_{j \geq 0} \bigcup_{T \in \mathbf{T}_j} T$$

where each  $\mathbf{T}_j$  is a collection of trees  $T$  each of which have energy bounded by  $C 2^{-j/2} |F|^{1/2}$ , density bounded by  $C 2^{-j/r'}$ , and satisfy

$$\sum_{T \in \mathbf{T}_j} |I_T| \lesssim 2^j.$$

For each  $j$  we apply Proposition 4.3 again, this time using (18) and (20) to write

$$\bigcup_{T \in \mathbf{T}_j} T = \bigcup_{k \geq 0} \bigcup_{T \in \mathbf{T}_{j,k}} T$$

where each tree  $T \in \mathbf{T}_{j,k}$  has energy bounded by  $C2^{-(j+k)/2}|F|^{1/2}$ , density bounded by  $C2^{-j/r'}$ , and satisfies

$$(58) \quad \sum_{T \in \mathbf{T}_{j,k}} |I_T| \lesssim 2^j.$$

Moreover, for every  $\ell \geq 0$

$$(59) \quad \left\| \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \right\|_{BMO} \lesssim 2^{2\ell} 2^{j+k} |F|^{-1}.$$

Inequality (58) implies  $\left\| \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \right\|_{L^1} \lesssim 2^{\ell+j}$ , and we may interpolate the  $L^1$  and the  $BMO$  bound. Here we use a standard technique involving the sharp maximal function from §5 in [11]. It follows that for  $1 \leq q < \infty$

$$\left\| \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \right\|_{L^q} \lesssim 2^{j+k+2\ell} |F|^{-1/q'}.$$

Let  $\epsilon > 0$  be small and  $C' > 0$  be large, depending on  $p, q, r$ . For each  $j, k, \ell$  define

$$G_{j,k,\ell} = \left\{ x : \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \geq C' |F|^{-1/q'} 2^{(1+\epsilon)(j+k+2\ell)} \right\}.$$

By Chebyshev's inequality, we have

$$|G_{j,k,\ell}| \leq c' 2^{-\epsilon(j+k+2\ell)},$$

so setting  $G = \bigcup_{j,k,\ell \geq 0} G_{j,k,\ell}$  we have  $|G| \leq 1/4$ .

Applying Minkowski's inequality gives

$$\begin{aligned} & \left\| \mathbb{1}_E \sum_{P \in \mathbf{P}} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G)} \\ & \leq \sum_{j,k \geq 0} \left( \left\| \mathbb{1}_E \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{I_T} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G_{j,k,0})} \right. \\ & \quad \left. + \sum_{\ell \geq 1} \left\| \mathbb{1}_E \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T \setminus 2^{\ell-1} I_T} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G_{j,k,\ell})} \right). \end{aligned}$$

From Hölder's inequality, Fubini's theorem, and the definition of  $G_{j,k,\ell}$ , it follows that the right side above is  $\lesssim (S_1 + S_2)$  where

$$S_1 = \sum_{j,k \geq 0} |F|^{-1/(q'r)} 2^{(1+\epsilon)(j+k)/r} \left( \sum_{T \in \mathbf{T}_{j,k}} \left\| \mathbb{1}_E \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^{r'}(\mathbb{R})}^{r'} \right)^{1/r'}$$

and

$$S_2 = \sum_{\substack{j,k \geq 0 \\ \ell \geq 1}} |F|^{-1/(q'r)} 2^{(1+\epsilon)(j+k+2\ell)/r} \left( \sum_{T \in \mathbf{T}_{j,k}} \left\| \mathbb{1}_E \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^{r'}(\mathbb{R} \setminus 2^{\ell-1} I_T)}^{r'} \right)^{1/r'}$$

Applying Proposition 5.1 with the energy and density bounds for trees  $T \in \mathbf{T}_{j,k}$ , we see that

$$\begin{aligned} S_2 &\lesssim \sum_{\substack{j,k \geq 0 \\ \ell \geq 1}} |F|^{-1/(q'r)} 2^{(1+\epsilon)(j+k+2\ell)/r} 2^{-\ell(N-10)} 2^{-(j+k)/2} |F|^{1/2} 2^{-j/r'} \left( \sum_{T \in \mathbf{T}_{j,k}} |I_T| \right)^{1/r'} \\ &\lesssim \sum_{\substack{j,k \geq 0 \\ \ell \geq 1}} 2^{(j+k)((1+\epsilon)(2/r)-1)/2} 2^{-\ell(N-14)} |F|^{1/2-1/(q'r)}. \end{aligned}$$

Choosing  $\epsilon$  small enough and  $q$  large enough so that

$$(1+\epsilon)\frac{2}{r} - 1 < 0 \text{ and } \frac{1}{2} - \frac{1}{q'r} < \frac{1}{p}$$

we have  $S_2 \lesssim |F|^{1/p}$ . We similarly obtain  $S_1 \lesssim |F|^{1/p}$ , thus giving (57).

We will finish by proving (57) for  $|F| \leq 1$ . Here, we let

$$G = \{x : \mathcal{M}[\mathbb{1}_F](x) > C''|F|\}$$

where  $\mathcal{M}$  is the Hardy-Littlewood maximal operator and  $C''$  is chosen large enough so that the weak-type 1-1 estimate for  $\mathcal{M}$  guarantees  $|G| \leq 1/4$ .

By Proposition 6.2 and the fact that  $p \geq r'$ , it will remain to show that

$$(60) \quad \left\| \mathbb{1}_E \sum_{P \in \mathbf{P}'} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G)} \lesssim |F|^{1/p}.$$

where  $\mathbf{P}' = \{P \in \mathbf{P} : I \not\subset G\}$ .

Finally, it follows from Proposition 6.1 that the energy of  $\mathbf{P}'$  is bounded above by  $C|F|$ . Repeatedly applying Propositions 4.3 and 4.4 we write  $\mathbf{P}'$  as the disjoint union

$$\mathbf{P}' = \bigcup_{j \geq 0} \bigcup_{T \in \mathbf{T}_j} T$$

where each  $\mathbf{T}_j$  is a collection of trees  $T$  each of which have energy bounded by  $C_0 2^{-j/2} |F|^{1/2}$ , density bounded by  $C_0 2^{-j/r'}$ , and satisfy

$$\sum_{T \in \mathbf{T}_j} |I_T| \lesssim 2^j.$$

We then have

$$\left\| \mathbb{1}_E \sum_{P \in \mathbf{P}'} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1} \leq \sum_{j \geq 0} \sum_{T \in \mathbf{T}_j} \left\| \mathbb{1}_E \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1}.$$

Applying Proposition 5.1, we see that the right side above is

$$\lesssim \sum_{j \geq 0} \sum_{T \in \mathbf{T}_j} \min(2^{-j/2} |F|^{1/2}, |F|) 2^{-j/r'} |I_T| \lesssim \sum_{j \geq 0} 2^{j/r} \min(2^{-j/2} |F|^{1/2}, |F|).$$

Summing over  $j$ , we see that the right side above is  $\lesssim |F|^{1/r'}$ . This finishes the proof, since  $p \geq r'$ .

## A. TRANSFERENCE

In this section we show how to obtain Theorem 1.1 from Theorem 1.2. We employ arguments from chapter VII in the monograph by Stein and Weiss [31] in their proof of De Leeuw's transference result ([21]). The following limiting relation is used:

**Lemma A.1.** ([31], p. 261). *Let  $m \in L^\infty(\mathbb{R}^d)$  with the property that every  $k \in \mathbb{Z}^d$  is a Lebesgue point of  $m$ . Define a convolution operator  $T$  on  $L^2(\mathbb{R}^d)$  by the Fourier transform identity  $\widehat{Tf}(\xi) = m(\xi)\widehat{f}(\xi)$  and a convolution operator on  $L^2(\mathbb{T}^d)$  by the relation  $[\widehat{Tf}]_k = m(k)f_k$  for the Fourier coefficients. Let, for  $x \in \mathbb{R}$ ,*

$$w(x) = e^{-\pi x^2} \text{ and } w_R(x) = R^{-d}w(x/R).$$

*Then, for all trigonometric polynomials  $P$  and  $Q$  (extended as 1-periodic functions in every variable) we have*

$$(61) \quad \int_{[0,1]^d} \widehat{T}[P](x)Q(x) dx = \lim_{R \rightarrow \infty} \int_{\mathbb{R}^d} T[Pw_{R/p'}](x)Q(x)w_{R/p}(x) dx.$$

We need also need the following elementary fact on Lorentz spaces.

**Lemma A.2.** *Let  $f \in L^{p,q}(\mathbb{T}^d)$  and extend  $f$  to a function  $f^{\text{per}}$  on  $\mathbb{R}^d$  which is 1-periodic in every variable. Let  $L > d/p$  and let  $w$  be a measurable function satisfying  $|w(x)| \leq (1 + |x|)^{-L}$ . Let  $w_R(x) = R^{-d}w(R^{-1}x)$ . Then*

$$\sup_{R \geq 1} \|f^{\text{per}}w_R\|_{L^{p,q}(\mathbb{R}^d)} \leq C_{p,q} \|f\|_{L^{p,q}(\mathbb{T}^d)}$$

*Proof.* We first assume  $p = q$ . Let  $Q_0 := [-\frac{1}{2}, \frac{1}{2}]^d$ . If  $N \in \mathbb{N}$  and  $N \leq R \leq N+1$  then  $w_R(x) \approx N^{-d}(1 + |n|/N)^{-Lp}$  for  $x \in n + Q_0$  and by the periodicity we can estimate  $\|f^{\text{per}}w_R\|_p^p$  by  $C \sum_{n \in \mathbb{Z}^d} N^{-d}(1 + |n|/N)^{-Lp} \|f\|_{L^p(Q_0)}^p$  which is  $\lesssim \|f\|_{L^p(Q_0)}^p$  since  $Lp > d$ . For fixed  $L$  we apply real interpolation in the range  $p < L/d$  and obtain the Lorentz space result.  $\square$

*Proof that Theorem 1.2 implies Theorem 1.1.* We shall assume that  $1 < p < \infty$ ,  $1 \leq q_1 < \infty$ , and  $q_1 \leq q_2 \leq \infty$  and prove that the  $L^{p,q_1}(\mathbb{R}) \rightarrow L^{p,q_2}(\mathbb{R}; V^r)$  for the partial sum operator  $\mathcal{S}$  on the real line implies the corresponding result on the torus, i.e.

$$(62) \quad \left\| \sup_K \sup_{0 \leq n_1 \leq \dots \leq n_K} \left( \sum_{i=1}^{K-1} |S_{n_{i+1}}f - S_{n_i}f|^r \right)^{1/r} \right\|_{L^{p,q_2}(\mathbb{T})} \lesssim \|f\|_{L^{p,q_1}(\mathbb{T})}$$

By two applications of the monotone convergence theorem it suffices to show for fixed  $K \in \mathbb{N}$  with  $K \geq 2$ , and fixed  $M \in \mathbb{N}$  that

$$(63) \quad \left\| \sup_{0 \leq n_1 \leq \dots \leq n_K \leq M} \left( \sum_{i=1}^{K-1} |S_{n_{i+1}}f - S_{n_i}f|^r \right)^{1/r} \right\|_{L^{p,q_2}(\mathbb{T})} \leq C \|f\|_{L^{p,q_1}(\mathbb{T})}$$

where  $C$  does not depend on  $M$  and  $K$ .

For  $\vec{n} = (n_1, \dots, n_K) \in \mathbb{N}_0^K$ ,  $1 \leq i \leq K-1$  define  $\mathcal{T}f(x, \vec{n}, i) = S_{n_{i+1}}f(x) - S_{n_i}f(x)$  if  $n_1 \leq \dots \leq n_K$  and  $\mathcal{T}f(x, \vec{n}, i) = 0$  otherwise. Then the inequality (63) just says that  $\mathcal{T}$  is bounded from  $L^{p,q_1}$  to  $L^{p,q_2}(\ell^\infty(\ell^r))$  where the  $\ell^\infty$  norm is taken for functions on the



finite set  $\{1, \dots, M\}^K$  and the  $\ell^r$  norm is for functions on  $\{1, \dots, K-1\}$ . By duality (63) follows from the  $L^{p', q'_2}(\ell^1(\ell^{r'})) \rightarrow L^{p', q'_1}$  inequality for the adjoint operator  $\mathcal{T}^*$ , i.e. from the inequality

$$(64) \quad \left| \int_0^1 \sum_{0 \leq n_1 \leq \dots \leq n_K \leq M} \sum_{i=1}^{K-1} [S_{n_{i+1}} f_{\vec{n}, i}(x) - S_{n_i} f_{\vec{n}, i}(x)] Q(x) dx \right| \\ \lesssim \left\| \sum_{\vec{n}} \left( \sum_{i=1}^{K-1} |f_{\vec{n}, i}|^{r'} \right)^{1/r'} \right\|_{L^{p', q'_2}(\mathbb{T})} \|Q\|_{L^{p, q_1}(\mathbb{T})}.$$

We fix an irrational number  $\lambda$  in  $(0, 1)$ , say  $\lambda = 1/\sqrt{2}$ . We then define “partial sum operators” for Fourier integrals by  $\widehat{\mathfrak{S}_t f}(\xi) = \chi_{[-\lambda, \lambda]}(\xi/t) \widehat{f}(\xi)$  and a corresponding partial sum operator  $\widetilde{\mathfrak{S}}_t$  on Fourier series by letting the  $k$ th Fourier coefficient of  $\widetilde{\mathfrak{S}}_t f$  be equal to  $\chi_{[-\lambda, \lambda]}(k/t) \widehat{f}_k$ . We define a function  $\nu : \mathbb{N}_0 \rightarrow \mathbb{N}_0$  as follows: set  $\nu(0) = 0$  and for  $n > 0$  let  $\nu(n)$  be the smallest positive integer  $\nu$  for which  $\lambda\nu > n$ . Notice that then

$$(65) \quad S_{n_{i+1}} f_{\vec{n}, i}(x) - S_{n_i} f_{\vec{n}, i}(x) = \widetilde{\mathfrak{S}}_{\nu(n_{i+1})} f_{\vec{n}, i} - \widetilde{\mathfrak{S}}_{\nu(n_i)} f_{\vec{n}, i}.$$

Now in order to prove (64) it clearly suffices to verify it for the case that the function  $f_{\vec{n}, i}$  and  $Q$  are trigonometric polynomials. The multipliers corresponding to  $\mathfrak{S}_t$  are continuous at every integer. Thus by (61) (applied with  $d = 1$ ) and (65) we see that (64) is implied by

$$(66) \quad \left| \sum_{0 \leq n_1 \leq \dots \leq n_K \leq M} \sum_{i=1}^{K-1} \int_0^1 [\mathfrak{S}_{\nu(n_{i+1})}(f_{\vec{n}, i} w_{R/p'}) - \mathfrak{S}_{\nu(n_i)}(f_{\vec{n}, i} w_{R/p'})] Q w_{R/p} dx \right| \\ \lesssim \left\| \sum_{\vec{n}} \left( \sum_{i=1}^{K-1} |f_{\vec{n}, i}|^{r'} \right)^{1/r'} \right\|_{L^{p', q'_2}(\mathbb{T})} \|Q\|_{L^{p, q_1}(\mathbb{T})}.$$

for sufficiently large  $R$ .

Now notice that  $\mathfrak{S}_t = \mathcal{S}_{\lambda t} - \mathcal{S}_{-\lambda t}$  so that the assumed  $L^{p, q_1}(\mathbb{R}) \rightarrow L^{p, q_2}(V^r, \mathbb{R})$  boundedness for the family  $\{\mathcal{S}_t\}$  implies the analogous statement for the family  $\{\mathfrak{S}_t\}$ . We run the duality argument in the reverse direction (now for functions defined on  $\mathbb{R}$ ) and deduce

$$\left| \int_{\mathbb{R}} \sum_{0 \leq n_1 \leq \dots \leq n_K \leq M} \sum_{i=1}^{K-1} (\mathfrak{S}_{\nu(n_{i+1})}[f_{\vec{n}, i} w_{R/p'}] - \mathfrak{S}_{\nu(n_i)}[f_{\vec{n}, i} w_{R/p'}]) Q w_{R/p} dx \right| \\ \lesssim \left\| \sum_{\vec{n}} \left( \sum_{i=1}^{K-1} |f_{\vec{n}, i}|^{r'} \right)^{1/r'} w_{R/p'} \right\|_{L^{p', q'_2}(\mathbb{R})} \|Q w_{R/p}\|_{L^{p, q_1}(\mathbb{R})}.$$

By Lemma A.2 the right hand side of this inequality is for  $R \geq \max\{p, p'\}$  bounded by the right hand side of (66). Thus we have established inequality (66) and this concludes the proof.  $\square$

## B. A VARIATIONAL MENSHOV-PALEY-ZYGMUND THEOREM

For  $\xi, x \in \mathbb{R}$  let

$$\mathcal{C}[f](\xi, x) = \int_{-\infty}^x e^{-2\pi i \xi x'} f(x') dx'.$$

Menshov, Paley, and Zygmund extended the Hausdorff-Young inequality by proving a version of the bound

$$(67) \quad \|\mathcal{C}[f]\|_{L_{\xi}^{p'}(L_x^{\infty})} \leq C_p \|f\|_{L^p(\mathbb{R})}$$

for  $1 \leq p < 2$ . The bound at  $p = 2$  is a special case of the much more difficult maximal inequality for the partial sum operator of proved by Carleson and Hunt. Interpolating the variational version, Theorem 1.2, at  $p = 2$  with a trivial estimate at  $p = 1$ , one sees that (67) may be strengthened to the bound

$$(68) \quad \|\mathcal{C}[f]\|_{L_{\xi}^{p'}(V_x^r)} \leq C_{p,r} \|f\|_{L^p(\mathbb{R})}$$

for  $1 \leq p \leq 2$  and  $r > p$ . It follows from the same arguments given in Section 2 that this range of  $r$  is the best possible. Our interest in this variational bound primarily stems from the fact, which will be proven in Appendix C, that it may be transferred, when  $r < 2$ , to give a corresponding estimate for certain nonlinear Fourier summation operators. The purpose of the present appendix is to give an easier alternate proof of (68) when  $p < 2$ . Note that Pisier and Xu [27] have proved closely related  $L^p \rightarrow L^2(V^p)$  inequalities for orthonormal systems of (not necessarily bounded) functions on an arbitrary measure space.

A now-famous lemma of Christ and Kiselev [5] asserts that if an integral operator

$$Tf(x) = \int_{\mathbb{R}} K(x, y) f(y) dy$$

is bounded from  $L^p(\mathbb{R})$  to  $L^q(X)$  for some measure space  $X$  and some  $q > p$ , thus

$$\|Tf\|_{L^q(X)} \leq A \|f\|_{L^p(\mathbb{R})},$$

then automatically the maximal function

$$T_*f(x) = \sup_{N \in \mathbb{R}} \left| \int_{y < N} K(x, y) f(y) dy \right|$$

is also bounded from  $L^p(\mathbb{R})$  to  $L^q(X)$ , with a slightly larger constant. Another way to phrase this is as follows. If we define the partial integrals

$$T_{\leq}f(x, N) = \int_{y < N} K(x, y) f(y) dy$$

then we have

$$(69) \quad \|T_{\leq}f\|_{L_x^q(L_N^{\infty})} \leq C_{p,q} A \|f\|_{L^p(\mathbb{R})}.$$

As was observed by Christ and Kiselev, this may be applied in conjunction with the Hausdorff-Young inequality to obtain (67) for  $p < 2$ .

The  $L_N^{\infty}$  norm can also be interpreted as the  $V_N^{\infty}$  norm, and we will now see that  $V^{\infty}$  can be replaced by  $V^r$  for  $r > p$ , thus giving (68) from the Hausdorff-Young inequality.

**Lemma B.1.** *Under the same assumptions, we have*

$$\|T_{\leq} f\|_{L_x^q(V_N^r)} \leq C_{p,q,r} A \|f\|_{L^p(\mathbb{R})}$$

for any  $r > p$ .

*Proof.* This follows by an adaption of the argument by Christ and Kiselev, or by the following argument. Without loss of generality we may take  $r < q$ , in particular  $r < \infty$ . We use a bootstrap argument. Let us make the *a priori* assumption that

$$(70) \quad \|T_{\leq} f\|_{L_x^q(V_N^r)} \leq BA \|f\|_{L^p(\mathbb{R})}$$

for some constant  $0 < B < \infty$ ; this can be accomplished for instance by truncating the kernel  $K$  appropriately. We will show that this *a priori* bound automatically implies the bound

$$(71) \quad \|T_{\leq} f\|_{L_x^q(V_N^r)} \leq (2^{1/r-1/p} BA + C_{p,q,r} A) \|f\|_{L^p(\mathbb{R})}$$

for some  $C_{p,q,r} > 0$ . This implies that the best bound  $B$  in the above inequality will necessarily obey the inequality

$$B \leq 2^{1/r-1/p} B + C_{p,q,r};$$

since  $r > p$ , this implies  $B \leq C'_{p,q,r}$  for some finite  $C'_{p,q,r}$ , and the claim follows.

It remains to deduce (71) from (70). Fix  $f$ ; we may normalize  $\|f\|_{L^p(\mathbb{R})} = 1$ . We find a partition point  $N_0$  in the real line which halves the  $L^p$  norm of  $f$ :

$$\int_{-\infty}^{N_0} |f(y)|^p dy = \int_{N_0}^{+\infty} |f(y)|^p dy = \frac{1}{2}.$$

Write  $f_-(y) = f(y) \mathbb{1}_{(-\infty, N_0]}(y)$  and  $f_+(y) = f(y) \mathbb{1}_{[N_0, +\infty)}(y)$ , thus  $\|f_-\|_{L^p(\mathbb{R})} = 2^{-1/p}$  and  $\|f_+\|_{L^p(\mathbb{R})} = 2^{-1/p}$ . We observe that

$$T_{\leq} f(x, N) = \begin{cases} T_{\leq} f_-(x, N) & \text{when } N \leq N_0 \\ T_{\leq} f_-(x) + T_{\leq} f_+(x, N) & \text{when } N > N_0 \end{cases}$$

Furthermore,  $T_{\leq} f_-(x, \cdot)$  and  $T_{\leq} f_+(x, \cdot)$  are bounded in  $L^\infty$  norm by  $O(T_* f(x))$ . Thus we have

$$\|T_{\leq} f(x, \cdot)\|_{V_N^r} \leq (\|T_{\leq} f_-(x, \cdot)\|_{V_N^r}^r + \|T_{\leq} f_+(x, \cdot)\|_{V_N^r}^r)^{1/r} + O(T_* f(x)).$$

(The  $O(T_* f(x))$  error comes because the partition used to define  $\|T_{\leq} f(x, \cdot)\|_{V_N^r}$  may have one interval which straddles  $N_0$ ). We take  $L^q$  norms of both sides to obtain

$$\|T_{\leq} f\|_{L_x^q V_N^r} \leq \|(\|T_{\leq} f_-(x, \cdot)\|_{V_N^r}^r + \|T_{\leq} f_+(x, \cdot)\|_{V_N^r}^r)^{1/r}\|_{L_x^q} + O(\|T_* f\|_{L_x^q}).$$

The error term is at most  $C_{p,q} A$  by the ordinary Christ-Kiselev lemma. For the main term, we take advantage of the fact that  $r < q$  to interchange the  $l^r$  and  $L^q$  norms, thus obtaining

$$\|T_{\leq} f\|_{L_x^q V_N^r} \leq (\|T_{\leq} f_-\|_{L_x^q V_N^r}^r + \|T_{\leq} f_+\|_{L_x^q V_N^r}^r)^{1/r} + O(C_{p,q} A).$$

By inductive hypothesis we thus have

$$\|T_{\leq} f\|_{L_x^q V_N^r} \leq ((2^{-1/p} BA)^r + (2^{-1/p} BA)^r)^{1/r} + O(C_{p,q} A),$$

and the claim follows. □

## C. VARIATION NORMS ON LIE GROUPS

In this appendix, we will show that certain  $r$ -variation norms for curves on Lie groups can be controlled by the corresponding variation norms of their “traces” on the Lie algebra as long as  $r < 2$ . This follows from work of Terry Lyons [23]; we present a self contained proof in this appendix. Combining this fact with the variational Menshov-Paley-Zygmund theorem of Appendix B, we rederive the Christ-Kiselev theorem on the pointwise convergence of the nonlinear Fourier summation operator for  $L^p(\mathbb{R})$  functions,  $1 \leq p < 2$ .

Let  $G$  be a connected finite-dimensional Lie group with Lie algebra  $\mathfrak{g}$ . We give  $\mathfrak{g}$  any norm  $\|\cdot\|_{\mathfrak{g}}$ , and push forward this norm using left multiplication by the Lie group to define a norm  $\|x\|_{T_g G} = \|g^{-1}x\|_{\mathfrak{g}}$  on each tangent space  $T_g G$  of the group. Observe that this norm structure is preserved under left group multiplication.

We can now define the *length*  $|\gamma|$  of a continuously differentiable path  $\gamma : [a, b] \rightarrow G$  by the usual formula

$$|\gamma| = \int_a^b \|\gamma'(t)\|_{T_{\gamma(t)} G} dt.$$

Observe that this notion of length is invariant under left group multiplication, and also under reparameterization of the path  $\gamma$ .

From this notion of length, we can define a metric  $d(g, g')$  on  $G$  as

$$d(g, g') = \inf_{\gamma: \gamma(a)=g, \gamma(b)=g'} |\gamma|$$

where  $\gamma$  ranges over all differentiable paths from  $g$  to  $g'$ . It is easy to see that this does indeed give a metric on  $G$ .

Given any continuous path  $\gamma : [a, b] \rightarrow G$  and  $1 \leq r < \infty$ , we define the  *$r$ -variation*  $\|\gamma\|_{V^r}$  of  $\gamma$  to be the quantity

$$\|\gamma\|_{V^r} = \sup_{a=t_0 < t_1 < \dots < t_n=b} \left( \sum_{j=0}^{n-1} d(\gamma(t_{j+1}), \gamma(t_j))^r \right)^{1/r}$$

where the infimum ranges over all partitions of  $[a, b]$  by finitely many times  $a = t_0, t_1, \dots, t_n = b$ . We can extend this to the  $r = \infty$  case in the usual manner as

$$\|\gamma\|_{V^\infty} = \sup_{a=t_0 < t_1 < \dots < t_n=b} \sup_{0 \leq j \leq n-1} d(\gamma(t_{j+1}), \gamma(t_j)),$$

and indeed it is clear that the  $V^\infty$  norm of  $\gamma$  is simply the diameter of the range of  $\gamma$ . The  $V^1$  norm of  $\gamma$  is finite precisely when  $\gamma$  is rectifiable, and when  $\gamma$  is differentiable it corresponds exactly with the length  $|\gamma|$  of  $\gamma$  defined earlier. It is easy to see the monotonicity property

$$\|\gamma\|_{V^p} \leq \|\gamma\|_{V^r} \text{ whenever } 1 \leq r \leq p \leq \infty$$

and the triangle inequalities

$$(\|\gamma_1\|_{V^r}^r + \|\gamma_2\|_{V^r}^r)^{1/r} \leq \|\gamma_1 + \gamma_2\|_{V^r} \leq \|\gamma_1\|_{V^r} + \|\gamma_2\|_{V^r}$$

where  $\gamma_1 + \gamma_2$  is the concatenation of  $\gamma_1$  and  $\gamma_2$ . A key fact about the  $V^r$  norms is that they can be subdivided:

**Lemma C.1.** *Let  $\gamma : [a, b] \rightarrow G$  be a continuously differentiable curve with finite  $V^r$  norm. Then there exists a decomposition  $\gamma = \gamma_1 + \gamma_2$  of the curve into two sub-curves such that*

$$\|\gamma_1\|_{V^r}, \|\gamma_2\|_{V^r} \leq 2^{-1/r} \|\gamma\|_{V^r}.$$

*Proof.* Let  $t_* = \sup\{t \in [a, b] : \|\gamma|_{[a,t]}\|_{V^r} \leq 2^{-1/r} \|\gamma\|_{V^r}\}$ . Letting  $\gamma_1 = \gamma|_{[a,t_*]}$  we have  $\|\gamma_1\|_{V^r} = 2^{-1/r} \|\gamma\|_{V^r}$ . The bound for  $\gamma_2 = \gamma|_{[t_*,b]}$  follows from the left triangle inequality above.  $\square$

Given a continuously differentiable curve  $\gamma : [a, b] \rightarrow G$ , we can define its *left trace*  $\gamma_l : [a, b] \rightarrow \mathfrak{G}$  by the formula

$$\gamma_l(t) = \int_a^t \gamma(s)^{-1} \gamma'(s) ds$$

Note that the trace is also a continuously differentiable curve, but taking values now in the Lie algebra  $\mathfrak{g}$  instead of  $G$ . Clearly  $\gamma_l$  is determined uniquely from  $\gamma$ . The converse is also true after specifying the initial point  $\gamma(a)$  of  $\gamma$ , since  $\gamma$  can then be recovered by solving the ordinary differential equation

$$(72) \quad \gamma'(t) = \gamma(t) \gamma_l'(t).$$

This equation is fundamental in the theory of eigenfunctions of a one-dimensional Schrödinger or Dirac operator, or equivalently in the study of the nonlinear Fourier transform; see, for example, [32], [25] for a full discussion. Basically for a fixed potential  $f(t)$  and a frequency  $k$ , the nonlinear Fourier transform traces out a curve  $\gamma(t)$  (depending on  $k$ ) taking values in a Lie group (e.g.  $SU(1, 1)$ ), and the corresponding left trace is essentially the ordinary linear Fourier transform.

It is easy to see that these curves have the same length (i.e. they have the same  $V^1$  norm):

$$(73) \quad |\gamma| = |\gamma_l|.$$

We now show that something similar is true for the  $V^r$  norms provided that  $r < 2$ .

**Lemma C.2.** *Let  $1 \leq r < 2$ , let  $G$  be a connected finite-dimensional Lie group, and let  $\|\cdot\|_{\mathfrak{g}}$  be a norm on the Lie algebra of  $G$ . Then there exist a constant  $C > 0$  depending only on these above quantities, such that for all smooth curves  $\gamma : [a, b] \rightarrow G$ , we have*

$$(74) \quad \|\gamma\|_{V^r} \leq \|\gamma_l\|_{V^r} + C \min(\|\gamma_l\|_{V^r}^2, \|\gamma_l\|_{V^r}^r)$$

and

$$(75) \quad \|\gamma_l\|_{V^r} \leq \|\gamma\|_{V^r} + C \min(\|\gamma\|_{V^r}^2, \|\gamma\|_{V^r}^r).$$

An analogous result holds for the *right trace*,  $\int_a^t \gamma'(s) \gamma(s)^{-1} ds$ , once the left-invariant norm on  $T_g G$  is replaced by a right-invariant norm.

*Proof.* We may take  $r > 1$  since the claim is already known for  $r = 1$  thanks to (73).

It shall suffice to prove the existence of a small  $\delta > 0$  such that we have the estimate

$$(76) \quad \|\gamma\|_{V^r} = \|\gamma_l\|_{V^r} + O(\|\gamma_l\|_{V^r}^2)$$

whenever  $\|\gamma_l\|_{V^r} \leq \delta$ , and similarly

$$(77) \quad \|\gamma_l\|_{V^r} = \|\gamma\|_{V^r} + O(\|\gamma\|_{V^r}^2)$$

whenever  $\|\gamma\|_{V^r} \leq \delta$ . (We allow the  $O()$  constants here to depend on  $r$ , the Lie group  $G$ , and the norm structure, but not on  $\delta$ ). Let us now see why these estimates will prove the lemma. Let us begin by showing that (76) implies (74). Certainly this will be the case if  $\gamma_l$  has  $V^r$  norm less than  $\delta$ . If instead  $\gamma_l$  has  $V^r$  norm larger than  $\delta$ , we can use Lemma C.1 repeatedly to partition it into  $O(\delta^{-r}\|\gamma_l\|_{V^r}^r)$  curves, all of whose  $V^r$  norms are less than  $\delta$ . These curves are the left-traces of various components of  $\gamma$ , and thus by (76) these components have a  $V^r$  norm bounded by some quantity depending on  $\delta$ . Concatenating these components together (using the triangle inequality) we obtain the result. A similar argument allows one to deduce (75) from (77).

Next, we observe that to prove the two estimates (76), (77) it suffices to just prove one of the two, for instance (76), as this will also imply (77) for  $\|\gamma\|_{V^r}$  sufficiently small by the usual continuity argument (look at the set of times  $t$  for which the restriction of  $\gamma$  to  $[a, b]$  obeys a suitable version of (77), and use (76) to show that this set is both open and closed if  $\|\gamma\|_{V^r}$  is small enough).

It remains to prove (76) for  $\delta$  sufficiently small. We shall in fact prove the more precise statement

$$(78) \quad \|\log(\gamma(a)^{-1}\gamma(b)) - \gamma_l(b)\|_{\mathfrak{g}} \leq K\|\gamma_l\|_{V^r}^2$$

for some absolute constant  $K > 0$  (and for  $\delta$  sufficiently small), where  $\log$  is the inverse of the exponential map  $\exp : \mathfrak{g} \rightarrow G$ . Note that it follows from a continuity argument as in the previous paragraph that if  $\delta$  is sufficiently small then  $\gamma(b)^{-1}\gamma(a)$  is sufficiently close to the identity that the logarithm is well-defined. Let us now see why (78) implies (76). Applying the inequality to any segment  $[t_j, t_{j+1}]$  in  $[a, b]$  we see that

$$\|\log(\gamma(t_j)^{-1}\gamma(t_{j+1})) - (\gamma_l(t_{j+1}) - \gamma_l(t_j))\|_{\mathfrak{g}} \leq K\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r}^2$$

and hence (since  $\delta$  is small)

$$d(\gamma(t_{j+1}), \gamma(t_j)) = \|\gamma_l(t_{j+1}) - \gamma_l(t_j)\|_{\mathfrak{g}} + O(\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r}^2).$$

Estimating  $O(\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r}^2)$  crudely by  $\|\gamma_l\|_{V^r} O(\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r})$  and taking the  $\ell^r$  sum in the  $j$  index, we see that for any partition  $a = t_0 < \dots < t_n = b$  we have

$$\left(\sum_{j=0}^{n-1} d(\gamma(t_{j+1}), \gamma(t_j))^r\right)^{1/r} = \left(\sum_{j=0}^{n-1} \|\gamma_l(t_{j+1}) - \gamma_l(t_j)\|_{\mathfrak{g}}^r\right)^{1/r} + O(\|\gamma_l\|_{V^r}^2).$$

Taking suprema over all partitions we obtain the result.

It remains to prove (78) for some suitably large  $K$ . This we shall do by an induction on scale (or ‘‘Bellman function’’) argument. Let us fix the smooth curve  $\gamma$ . We shall prove the estimate for all subcurves of  $\gamma$ , i.e. for all intervals  $[t_1, t_2]$  in  $[a, b]$ , we shall prove that

$$(79) \quad \|\log(\gamma(t_1)^{-1}\gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_1))\|_{\mathfrak{g}} \leq K\|\gamma_l|_{[t_1, t_2]}\|_{V^r}^2.$$

Let us first prove this in the case when the interval  $[t_1, t_2]$  is sufficiently short, say of length at most  $\epsilon$  for some very small  $\epsilon$  (depending on  $\gamma$ ). In that case, we perform a Taylor expansion to obtain

$$(80) \quad \gamma_l(t) = \gamma_l(t_1) + \gamma_l'(t_1)(t - t_1) + \frac{1}{2}\gamma_l''(t_1)(t - t_1)^2 + O_\gamma((t - t_1)^3)$$

and

$$(81) \quad \gamma'_l(t) = \gamma'_l(t_1) + \gamma''_l(t_1)(t - t_1) + O_\gamma((t - t_1)^2)$$

when  $t \in [t_1, t_2]$ , and where the  $\gamma$  subscript in  $O_\gamma$  means that the constants here are allowed to depend on  $\gamma$  (more specifically, on the  $C^3$  norm of  $\gamma$ ), and the  $O()$  is with respect to the  $\|\cdot\|_{\mathfrak{g}}$  norm. Also we remark that as  $\gamma$  is assumed smooth,  $\gamma'_l(t_1)$  is bounded away from zero. It is then an easy matter to conclude that

$$(82) \quad \|\gamma_l|_{[t_1, t_2]}\|_{V^r} \geq \frac{1}{2} \|\gamma'_l(t_1)\|_{\mathfrak{g}} |t_2 - t_1|$$

if  $\epsilon$  is sufficiently small depending on  $\gamma$ . On the other hand, from (72) and (81) we have

$$\gamma'(t) = \gamma(t)(\gamma'_l(t_1) + \gamma''_l(t_1)(t - t_1) + O_\gamma((t - t_1)^2))$$

from which one may conclude that

$$\gamma(t) = \gamma(t_1) \exp(\gamma'_l(t_1)(t - t_1) + \frac{1}{2} \gamma''_l(t_1)(t - t_1)^2 + O(\|\gamma'_l(t_1)\|_{\mathfrak{g}}^2 |t - t_1|^2) + O_\gamma((t - t_1)^3))$$

for all  $t \in [t_1, t_2]$ , if  $\gamma$  is sufficiently small. We rewrite this as

$$\begin{aligned} & \log(\gamma(t_1)^{-1} \gamma(t)) \\ &= \gamma'_l(t_1)(t - t_1) + \frac{1}{2} \gamma''_l(t_1)(t - t_1)^2 + O(\|\gamma'_l(t_1)\|_{\mathfrak{g}}^2 |t - t_1|^2) + O_\gamma((t - t_1)^3), \end{aligned}$$

and then specialize to the case  $t = t_2$ . By (80), we have

$$\log(\gamma(t_1)^{-1} \gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_1)) = O(\|\gamma'_l(t_1)\|_{\mathfrak{g}}^2 |t_2 - t_1|^2) + O_\gamma((t_2 - t_1)^3),$$

and hence by (82) we have (79) if  $t_2 - t_1$  is small enough (depending on  $\gamma$ ) and  $K$  is large enough (*independent* of  $\gamma$ ).

This proves (79) when the interval  $[t_1, t_2]$  is small enough. By (82), it also proves (79) when  $\|\gamma_l|_{[t_1, t_2]}\|_{V^r}$  is sufficiently small. To conclude the proof of (79) in general, we now assert the following inductive claim: if (79) holds whenever  $\|\gamma_l|_{[t_1, t_2]}\|_{V^r} < \epsilon$  and some given  $0 < \epsilon \leq \delta$ , then it also holds whenever  $\|\gamma_l|_{[t_1, t_2]}\|_{V^r} < 2^{1/r} \epsilon$ , providing that  $K$  is sufficiently large (*independent* of  $\epsilon$ ) and  $\delta$  is sufficiently small (depending on  $K$ , but *independent* of  $\epsilon$ ). Iterating this we will obtain the claim (79) for all intervals  $[t_1, t_2]$  in  $[a, b]$ .

It remains to prove the inductive claim. Let  $[t_1, t_2]$  be any subinterval of  $[a, b]$  such that the quantity  $A = \|\gamma_l|_{[t_1, t_2]}\|_{V^r}$  is less than  $2^{1/r} \epsilon$ . Applying Lemma C.1, we may subdivide  $[t_1, t_2] = [t_1, t_*] \cup [t_*, t_2]$  such that

$$\|\gamma_l|_{[t_1, t_*]}\|_{V^r}, \|\gamma_l|_{[t_*, t_2]}\|_{V^r} \leq 2^{-1/r} A < \epsilon \leq r.$$

By the inductive hypothesis, we thus have

$$\|\log(\gamma(t_1)^{-1} \gamma(t_*)) - (\gamma_l(t_*) - \gamma_l(t_1))\|_{\mathfrak{g}} \leq K 2^{-2/r} A^2$$

and

$$\|\log(\gamma(t_*)^{-1} \gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_*))\|_{\mathfrak{g}} \leq K 2^{-2/r} A^2.$$

In particular, we have

$$\begin{aligned} \|\log(\gamma(t_1)^{-1} \gamma(t_*))\|_{\mathfrak{g}} &\leq \|\gamma_l(t_*) - \gamma_l(t_1)\|_{\mathfrak{g}} + K 2^{-2/r} A^2 \\ &\leq \|\gamma_l|_{[t_1, t_*]}\|_{V^r} + O(K A^2) \\ &= O(A(1 + KA)) = O(A(1 + K\delta)) = O(A) \end{aligned}$$



if  $\delta$  is sufficiently small depending on  $K$ . Similarly we have

$$\|\log(\gamma(t_*)^{-1}\gamma(t_2))\|_{\mathfrak{g}} = O(A)$$

and hence by the Baker-Campbell-Hausdorff formula (if  $\delta$  is sufficiently small)

$$\|\log(\gamma(t_1)^{-1}\gamma(t_2)) - \log(\gamma(t_1)^{-1}\gamma(t_*) - \log(\gamma(t_*)^{-1}\gamma(t_2))\|_{\mathfrak{g}} = O(A^2).$$

By the triangle inequality, we thus have

$$\|\log(\gamma(t_1)^{-1}\gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_1))\|_{\mathfrak{g}} \leq 2K2^{-2/r}A^2 + O(A^2).$$

We now use the hypothesis  $r < 2$ , which forces  $2 \times 2^{-2/r} < 1$ . If  $K$  is large enough (depending on  $r$ , but independently of  $\delta$ ,  $A$ , or  $\epsilon$ ) we thus have (79). This closes the inductive argument.  $\square$

Letting  $w, v$  be any elements of the Lie algebra  $\mathfrak{g}$ , one can define a nonlinear Fourier summation operator associated to  $G, w, v$  by means of the left trace

$$\mathcal{NC}[f](k, 0) = I$$

$$\frac{\partial}{\partial x} \mathcal{NC}[f](k, x) = \mathcal{NC}[f](k, x) \left( \operatorname{Re}(e^{-2\pi i k x} f(x))w + \operatorname{Im}(e^{-2\pi i k x} f(x))v \right)$$

or (giving a different operator) by the right trace

$$\mathcal{NC}[f](k, 0) = I$$

$$\frac{\partial}{\partial x} \mathcal{NC}[f](k, x) = \left( \operatorname{Re}(e^{-2\pi i k x} f(x))w + \operatorname{Im}(e^{-2\pi i k x} f(x))v \right) \mathcal{NC}[f](k, x).$$

Above,  $k, x \in \mathbb{R}$ ,  $\mathcal{NC}[f]$  takes values in  $G$ ,  $I$  is the identity element of  $G$ , and  $\operatorname{Re}, \operatorname{Im}$  are the real and imaginary parts of a complex number. An example of interest is given by  $G = SU(1, 1)$ , and

$$w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad v = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

Combining Lemma C.2 with the variational Menshov-Paley-Zygmund theorem of the previous section, we obtain a variational version of the Christ-Kiselev theorem [6]. Namely, we see that for  $1 \leq p < 2$  and  $r > p$

$$\|\mathbb{1}_{|\mathcal{NC}[f]| \leq 1} \mathcal{NC}[f]\|_{L_k^{p'}(V_x^r)} \leq C_{p,r,G,w,v} \|f\|_{L^p(\mathbb{R})}$$

and

$$\|\mathbb{1}_{|\mathcal{NC}[f]| \geq 1} \mathcal{NC}[f]\|_{L_k^{p'/r}(V_x^r)}^{1/r} \leq C_{p,r,G,w,v} \|f\|_{L^p(\mathbb{R})}.$$

Note that the usual logarithms are hidden in the  $d$  metric we have placed on the Lie group  $G$ .

Extending these estimates to the case  $p = 2$  is an interesting and challenging problem, even when  $r = \infty$ , which would corresponds to a nonlinear Carleson theorem. Lemma C.2 cannot be extended to any exponent  $r \geq 2$ . Sandy Davie and the fifth author of this paper have an unpublished example of a curve in the Lie group  $SU(1, 1)$  with trace in the subspace of  $\mathfrak{su}(1, 1)$  of matrices vanishing on the diagonal so that the diameter of the curve is not controlled by the 2-variation of the trace.

Terry Lyons' machinery [24] via iterated integrals faces an obstruction in a potential application to a nonlinear Carleson theorem because of the unboundedness results for the iterated integrals shown in [26].

## D. AN APPLICATION TO ERGODIC THEORY

Wiener-Wintner type theorems is an area in ergodic theory that is most closely related to the study of Carleson's operator. In [19], Lacey and Terwilliger prove the following singular integral variant of the Wiener-Wintner theorem:

**Theorem D.1.** *For  $1 < p$ , all measure preserving flows  $\{T_t : t \in \mathbb{R}\}$  on a probability space  $(X, \mu)$  and functions  $f \in L^p(\mu)$ , there is a set  $X_f \subset X$  of probability one, so that for all  $x \in X_f$  we have that the limit*

$$\lim_{s \rightarrow 0} \int_{s < |t| < 1/s} e^{i\theta t} f(T_t x) \frac{dt}{t} .$$

*exists for all  $\theta \in \mathbb{R}$ .*

One idea to approach such convergence results is to study quantitative estimates in the parameter  $s$  that imply convergence, as pioneered by Bourgain's paper [1] in similar context. We first need to pass to a mollified variant of the above theorem:

**Theorem D.2.** *Let  $\phi$  be a function on  $\mathbb{R}$  in the Wiener space, i.e. the Fourier transform  $\widehat{\phi}$  is in  $L^1(\mathbb{R})$ . For  $1 < p$ , all measure preserving flows  $\{T_t : t \in \mathbb{R}\}$  on a probability space  $(X, \mu)$  and functions  $f \in L^p(\mu)$ , there is a set  $X_f \subset X$  of probability one, so that for all  $x \in X_f$  we have that the limits*

$$\begin{aligned} \lim_{s \rightarrow \infty} \int e^{i\theta t} f(T_t x) \phi(st) \frac{dt}{t} , \\ \lim_{s \rightarrow 0} \int e^{i\theta t} f(T_t x) \phi(st) \frac{dt}{t} . \end{aligned}$$

*exist for all  $\theta \in \mathbb{R}$ .*

This theorem clearly follows from an a priori estimate

$$\left\| \sup_{\theta} \left\| \int e^{i\theta t} f(T_t x) \phi(st) \frac{dt}{t} \right\|_{V^r(s)} \right\|_{L^p(x)} \leq C \|f\|_{L^p}$$

for  $r > \max(2, p')$ . Here we have written  $V^r(s)$  for the variation norm taken in the parameter  $s$  of the expression inside, and likewise for  $L^p(x)$ . The variation norm is the strongest norm widely used in this context, while Lacey and Terwilliger use a weaker oscillation norm in the proof of their Theorem.

By a standard transfer method ([3], [7]) involving replacing  $f$  by translates  $T_y f$  and an averaging procedure in  $y$ , the a priori estimate can be deduced from an analogous estimate on the real line

$$(83) \quad \left\| \sup_{\xi} \left\| \int e^{\xi i t} f(x+t) \phi(st) \frac{dt}{t} \right\|_{V^r(s)} \right\|_{L^p(x)} \lesssim \|f\|_{L^p} .$$

The main purpose of this appendix is to show how this estimate (83) can be deduced from the main theorem of this paper by an averaging argument. We write the  $V^r(s)$  norm explicitly and expand  $\phi$  into a Fourier integral to obtain for the left hand side of (83) the expression

$$\left\| \sup_{\xi} \sup_{s_0 < s_1 < \dots < s_K} \left( \sum_{k=1}^K \left| \int \int e^{\xi i t} f(x+t) e^{i\eta(s_k - s_{k-1})t} \frac{dt}{t} \widehat{\phi}(\eta) d\eta \right|^r \right)^{1/r} \right\|_{L^p(x)} .$$

Now pulling the integral in  $\eta$  out of the various norms and considering only positive  $\eta$  (with the case of negative  $\eta$  being similar) and defining  $\xi_k = \xi + \eta s_k$  we obtain the upper bound

$$\int_{\eta>0} \left\| \sup_{\xi_0 < \xi_1 < \dots < \xi_K} \left( \sum_{k=1}^K \left| \int e^{i(\xi_k - \xi_{k-1})t} f(x+t) \frac{dt}{t} \right|^r \right)^{1/r} \right\|_{L^p(x)} |\widehat{\phi}(\eta)| d\eta .$$

Now applying the variational Carleson estimate and doing the trivial integral in  $\eta$  bounds this term by a constant times  $\|f\|_{L^p}$ .

*Remark D.3.* To prove the Lacey-Terwilleger theorem D.1 from the mollified version, one may approximate the characteristic functions used as cutoff functions by Wiener space functions so that the difference is small in  $L^1$  norm. Then at least for  $f$  in  $L^\infty$  one can show convergence of the limits by an approximation argument, even though one will not recover the full strength of the quantitative estimate in the Wiener space setting. The result for  $f$  in  $L^\infty$  can then be used as a dense subclass result in other  $L^p$  spaces, which can be handled by easier maximal function estimates and further approximation arguments.

*Remark D.4.* The classical version of the Wiener-Wintner theorem does not invoke singular integrals but more classical averages of the type

$$\frac{1}{2s} \int_{|t|<s} e^{i\theta t} f(T_t x) dt .$$

We note that the same technique as above may be applied to these easier averages.

## REFERENCES

- [1] Jean Bourgain, *Pointwise ergodic theorems for arithmetic sets*, Inst. Hautes Études Sci. Publ. Math. (1989), no. 69, 5–45. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.
- [2] James T. Campbell, Roger L. Jones, Karin Reinhold and Máté Wierdl, *Oscillation and variation for the Hilbert transform*, Duke Math. J., **105** (2000), 59–83.
- [3] Alberto P. Calderón, *Ergodic theory and translation-invariant operators*, Proc. Nat. Acad. Sci. U.S.A. **59** (1968) 349–353.
- [4] Lennart Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.
- [5] Michael Christ and Alexander Kiselev, *Maximal functions associated to filtrations*, J. Funct. Anal. **179** (2001), no. 2, 409–425.
- [6] ———, *WKB asymptotic behavior of almost all generalized eigenfunctions for one-dimensional Schrödinger operators with slowly decaying potentials*, J. Funct. Anal. **179** (2001), no. 2, 426–447.
- [7] Ronald R. Coifman and Guido Weiss, *Transference methods in analysis*. Conference Board of the Mathematical Sciences Regional Conference Series in Mathematics, No. 31. American Mathematical Society, Providence, R.I., 1976.
- [8] Ciprian Demeter, Michael T. Lacey, Terence Tao, and Christoph Thiele, *Breaking the duality in the return times theorem*, Duke Math. J. **143** (2008), no. 2, 281–355.
- [9] ———, *The Walsh model for  $M_2^*$  Carleson*, Rev. Mat. Iberoamericana **24** (2008), no. 3, 721–744.
- [10] Charles Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. (2) **98** (1973), 551–571.
- [11] Charles Fefferman and Elias M. Stein,  *$H^p$  spaces of several variables*. Acta Math. **129** (1972), no. 3–4, 137–193.
- [12] Loukas Grafakos, Terence Tao, and Erin Terwilleger,  *$L^p$  bounds for a maximal dyadic sum operator*, Math. Z. **246** (2004), no. 1–2, 321–337.
- [13] Richard A. Hunt, *On the convergence of Fourier series*, Orthogonal Expansions and their Continuous Analogues (Proc. Conf., Edwardsville, Ill., 1967), Southern Illinois Univ. Press, Carbondale, Ill., 1968, pp. 235–255.

- [14] Roger L. Jones, Robert Kaufman, Joseph Rosenblatt, and Máté Wierdl, *Oscillation in ergodic theory*, Erg. Th. & Dyn. Sys., **18** (1998), 889–936.
- [15] Roger L. Jones, Andreas Seeger, and James Wright, *Strong variational and jump inequalities in harmonic analysis*, Trans. Amer. Math. Soc. **360** (2008), no. 12, 6711–6742.
- [16] Roger L. Jones and Gang Wang, *Variation inequalities for the Fejér and Poisson kernels*, Trans. Amer. Math. Soc. **356** (2004), no. 11, 4493–4518.
- [17] Carlos E. Kenig and Peter A. Tomas, *Maximal operators defined by Fourier multipliers*, Studia Math. **68** (1980), no. 1, 79–83.
- [18] Michael Lacey, *Issues related to Rubio de Francia’s Littlewood-Paley inequality*, NYJM Monographs, **2**. State University of New York, University at Albany, Albany, NY, (2007) 36 pp. (electronic).
- [19] Michael Lacey and Erin Terwilleger, *A Wiener-Wintner theorem for the Hilbert transform*, Ark. Mat. **46** (2008), no. 2, 315–336.
- [20] Michael Lacey and Christoph Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. **7** (2000), no. 4, 361–370.
- [21] Karel de Leeuw, *On  $L_p$  multipliers*, Ann. of Math. **81** (1965), 364–379.
- [22] Dominique Lépingle, *La variation d’ordre  $p$  des semi-martingales*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **36** (1976), no. 4, 295–316.
- [23] Terry Lyons, *Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young*, Math. Res. Lett. **1** (1994), no. 4, 451–464.
- [24] ———, *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana **14** (1998), 215–310.
- [25] Camil Muscalu, Terence Tao, and Christoph Thiele, *A Carleson theorem for a Cantor group model of the scattering transform*, Nonlinearity **16** (2003), no. 1, 219–246.
- [26] ———, *A counterexample to a multilinear endpoint question of Christ and Kiselev*, Math. Res. Lett. **10** (2003) no. 2-3, 237–246.
- [27] Gilles Pisier and Quanhua Xu, *The strong  $p$ -variation of martingales and orthogonal series*, Prob. Theory Related Fields, **77** (1988), 497–451.
- [28] Jinghua Qian, *The  $p$ -variation of partial sum processes and the empirical process*, Ann. Probab. **26** (1998), no. 3, 1370–1383.
- [29] Tong Seng Quek, *Littlewood-Paley type inequality on  $\mathbb{R}$* , Math. Nachr. 248/249 (2003), 151–157.
- [30] José L. Rubio de Francia, *A Littlewood-Paley inequality for arbitrary intervals*, Revista Mat. Iberoamericana **1** (1985), no. 2, 1–14.
- [31] Elias M. Stein and Guido Weiss, *Introduction to Fourier analysis on Euclidean spaces*. Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971.
- [32] Terence Tao and Christoph Thiele, *Nonlinear Fourier analysis*, IAS/Park City Math. Ser., to appear.
- [33] Christoph Thiele, *Wave packet analysis*, CBMS Regional Conference Series in Mathematics, vol. 105, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2006.
- [34] Norbert Wiener, *The quadratic variation of a function and its Fourier coefficients*, MIT Journal of Math. and Physics **3**(1924) pp. 72–94.

R. OBERLIN, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095-1555, USA

*E-mail address:* oberlin@math.ucla.edu

A. SEEGER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN, 480 LINCOLN DRIVE, MADISON, WI, 53706, USA

*E-mail address:* seeger@math.wisc.edu

T. TAO, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095-1555, USA

*E-mail address:* tao@math.ucla.edu

C. THIELE, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095-1555, USA

*E-mail address:* `thiele@math.ucla.edu`

J. WRIGHT, SCHOOL OF MATHEMATICS, UNIVERSITY OF EDINBURGH, JCMB, KING'S BUILDINGS,  
MAYFIELD ROAD, EDINBURGH EH9 3JZ, SCOTLAND

*E-mail address:* `wright@maths.ed.ac.uk`

# A VARIATION NORM CARLESON THEOREM

RICHARD OBERLIN   ANDREAS SEEGER   TERENCE TAO  
CHRISTOPH THIELE   JAMES WRIGHT

ABSTRACT. We strengthen the Carleson-Hunt theorem by proving  $L^p$  estimates for the  $r$ -variation of the partial sum operators for Fourier series and integrals, for  $r > \max\{p', 2\}$ . Four appendices are concerned with transference, a variation norm Menshov-Paley-Zygmund theorem, and applications to nonlinear Fourier transforms and ergodic theory.

## 1. INTRODUCTION

For an integrable function  $f$  on the circle group  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ , and  $k \in \mathbb{Z}$  we denote by  $\widehat{f}_k = \int_0^1 f(y) e^{-2\pi i k y} dy$  the Fourier coefficients and consider the partial sum operators  $S_n$  for the Fourier series,

$$(1) \quad S_n f(x) \equiv S[f](n, x) = \sum_{k=-n}^n \widehat{f}_k e^{2\pi i k x};$$

here  $n \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ . The celebrated theorem by Carleson [4] states that if  $f$  is square integrable then  $S_n f$  converges to  $f$  almost everywhere. Hunt [13] extended this result to  $L^p(\mathbb{T})$  functions, for  $1 < p < \infty$ , and proved the inequality

$$(2) \quad \left\| \sup_n |S_n f| \right\|_{L^p(\mathbb{T})} \leq C \|f\|_{L^p(\mathbb{T})}$$

for all  $f \in L^p(\mathbb{T})$ ; see also [10], [20], and [12] for other proofs of this fact.

The purpose of this paper is to strengthen the Carleson-Hunt result for  $L^p$  functions,  $1 < p < \infty$ , and show that, for  $r > \max\{2, p'\}$ , the (strong)  $r$ -variation of the sequence  $\{S_n f(x)\}_{n \in \mathbb{N}_0}$  is finite for almost every  $x \in [0, 1]$ . This can be interpreted as a statement about the rate of convergence. To fix notation, we consider real or complex valued sequences  $\{a_n\}_{n \in \mathbb{N}_0}$  and define their  $r$ -variation to be

$$(3) \quad \|a\|_{V^r} = \sup_K \sup_{n_0 < \dots < n_K} \left( \sum_{\ell=1}^K |a_{n_\ell} - a_{n_{\ell-1}}|^r \right)^{1/r}$$

where the sup is taken over all  $K$  and then over all increasing sequences of nonnegative integers  $n_0 < \dots < n_K$ . Note that the variation norms are monotone decreasing in the parameter  $r$ . Next, for a sequence  $F = \{F_n\}$  of Lebesgue measurable functions one defines the  $r$ -variation of  $F$  at  $x$ , sometimes denoted by  $\mathcal{V}^r F(x)$  as the  $V^r$  norm of the

---

*Date:* August 26, 2010.

1991 *Mathematics Subject Classification.* 42B15.

R.O. partially supported by NSF VIGRE grant DMS 0502315. A.S. partially supported by NSF grant DMS 0652890. T.T. partially supported by NSF Research Award DMS-0649473, the NSF Waterman award and a grant from the McArthur Foundation. C.Th. partially supported by NSF grant DMS 0701302. J.W. partially supported by an EPSRC grant.

sequence  $\{F_n(x)\}$ . We denote the  $r$ -variation of the sequence  $F_n = S_n f$  by  $\mathcal{V}^r S[f]$ . The variation norms and the  $r$ -variation operator can be defined in a similar fashion, if the index set  $\mathbb{N}_0$  is replaced by another subset of  $\mathbb{R}$  (often  $\mathbb{R}^+$  or  $\mathbb{R}$  itself).

Let  $r' := r/(r-1)$ , the conjugate exponent of  $r$ .

**Theorem 1.1.** *Suppose  $r > 2$  and  $r' < p < \infty$ . Then, for every  $f \in L^p(\mathbb{T})$ ,*

$$(4) \quad \|\mathcal{V}^r S[f]\|_{L^p(V^r)} \leq C_{p,r} \|f\|_{L^p}.$$

*At the endpoint  $p = r'$  a restricted weak type result holds; namely, for any  $f \in L^{r',1}(\mathbb{T})$  the function  $\mathcal{V}^r S[f]$  belongs to  $L^{r',\infty}(\mathbb{T})$ .*

It is immediate that (4) for  $r < \infty$  implies a quantitative form of almost everywhere convergence of Fourier series, improving over the standard qualitative result utilizing the weaker  $r = \infty$  inequality and convergence on a dense subclass of functions.

As will be discussed in Section 2, the conditions on the exponents in (4) are sharp. Moreover, in the endpoint case  $p = r'$  the Lorentz space  $L^{r',\infty}$  cannot be replaced by a smaller Lorentz space.

By standard transference arguments (see Appendix A) Theorem 1.1 is implied by a result on the partial (inverse) Fourier integral of a Schwartz function  $f$  on  $\mathbb{R}$  is defined as

$$\mathcal{S}[f](\xi, x) = \int_{-\infty}^{\xi} \widehat{f}(\eta) e^{2\pi i \eta x} d\eta$$

where  $\widehat{f}(\eta) = \int f(y) e^{-2\pi i y \eta} dy$  defines the Fourier transform of  $f$ .

**Theorem 1.2.** *Suppose  $r > 2$ . Then  $\mathcal{S}$  extends to a bounded operator  $\mathcal{S} : L^p \rightarrow L^p(V^r)$  for  $r' < p < \infty$ . Moreover  $\mathcal{S}$  maps  $L^{r',1}$  boundedly to  $L^{r',\infty}(V^r)$ .*

Note that if in the above definition of the mixed  $L^p(V^r)$  spaces we interchange the order between integration in the  $x$  variable and taking the supremum over the choices of  $K$  and the points  $\xi_0$  to  $\xi_K$  so that these choices become independent of the variable  $x$ , then the estimates corresponding to Theorem 1.2 are weaker; they follow from a square function inequality of Rubio de Francia [31] for  $p \geq 2$ , see also [30] for a related endpoint result for  $p < 2$ , and [18] for a proof of Rubio de Francia's inequality which is closer to the methods of this paper.

While the concept of  $r$ -variation norm is at least as old as Wiener's 1920s paper on quadratic variation [35], variational estimates have been pioneered by D. Lépingle ([22]) who proved them for martingales. Simple proofs of Lépingle's result based on jump inequalities have been given by Pisier and Xu [27] and by Bourgain [1], and applications to other families of operators in harmonic analysis such as families of averages and singular integrals have been considered in [1], and the subsequent papers [14], [2], [15] (*cf.* the bibliography of [15] for more references). Bourgain [1] used variation norm estimates (or related oscillation estimates which are intermediate in difficulty between maximal and variation norm estimates) to prove pointwise convergence results without previous knowledge that pointwise convergence holds for a dense subclass of functions. Such dense subclasses of functions, while usually available in the setting of analysis on Euclidean space, are less abundant in the ergodic theory setting. In Appendix D we



demonstrate the use of Theorem 1.2 in the setting of Wiener-Wintner type theorems as developed in [19]. We note that the Carleson-Hunt theorem has previously been generalized by using other norms in place of the variation norm, see for example the use of oscillation norms in [19], and the  $M_2^*$  norms in [8], [9].

We are also motivated by the fact that variation norms are in certain situations more stable under nonlinear perturbation than supremum norms. For example one can deduce bounds for certain  $r$ -variational lengths of curves in Lie groups from the corresponding lengths of the “trace” of the curves in the corresponding Lie algebras, see Appendix C for definitions and details. What we have in mind is proving Carleson type theorems for nonlinear perturbations of the Fourier transform as discussed in [25], [26]. Unfortunately the naive approach fails and the ultimate goal remains unattained since we only know the correlation between lengths of the trace and the original curve for  $r < 2$ , while the variational Carleson theorem only holds for  $r > 2$ . Nonetheless, this method allows one to see that a variational version of the Christ-Kiselev theorem [6] follows from a variational Menshov-Paley-Zygmund theorem which we prove in Appendix B. The variational Carleson inequality can be viewed as an endpoint in this theory.

Our proof of Theorem 1.2 will follow the method of [20] as refined in [12]. Naturally one has to invoke variation norm results in the setting of individual trees, which is achieved by adapting D. Lépingle’s result ([22]) to the setting of a tree. The authors initially had a proof of the case  $p > 2$  and  $r > p$  of Theorem 1.2 more akin to [20], while improvement to  $r > 2$  for such  $p$  provided a stumbling block. This stumbling block was removed by better accounting for trees of given energy, as described in the remarks leading to Proposition 4.3. In Section 3 we reduce the problem to that of bounding certain model operators which map  $f$  to linear combinations of wave-packets associated to collections of multitiles. In Section 5 we bound the model operators when the collection of multitiles is of a certain type called a tree; this bound is in terms of two quantities, energy and density, which are associated to the tree. These quantities are defined in Section 4 and an algorithm is given to decompose an arbitrary collection of multitiles into a union of trees with controlled energy and density. Section 6 contains two auxiliary estimates. All these ingredients are combined to complete the proof in Section 7.

The authors would like to thank the anonymous referee for valuable comments on the submitted draft of the paper.

*Some notation.* For two quantities  $A$  and  $B$  let  $A \lesssim B$  denote the statement that  $A \leq CB$  for some constant  $C$  (possibly depending on the parameters  $p$  and  $r$ ). The Lebesgue measure of a set  $E$  is denoted either by  $|E|$  or by  $\text{meas}(E)$ . The indicator function of  $E$  is denoted by  $\mathbb{1}_E$ . For a subset  $E \subset \mathbb{R}$  and  $a \in \mathbb{R}$  we set  $a + E = E + a = \{x : x - a \in E\}$ . If  $I$  is a finite interval  $I$  with center  $c(I)$  we denote by  $CI$  the  $C$  dilate of  $I$  with respect to its center, i.e. the set of all  $x$  for which  $c(I) + \frac{x - c(I)}{C} \in I$ .

## 2. OPTIMALITY OF THE EXPONENTS

Since Theorem 1.2 implies Theorem 1.1 (*cf.* Appendix A) we have to discuss the optimality only for the Fourier series case. The necessity of the condition  $r > 2$  follows from a corresponding result for the Cesaro means; its proof by Jones and Wang [16] was based on a probabilistic result of Qian [29].



We show the necessity of the condition  $p > r'$  in Theorem 1.1. Let

$$D_n(x) = \sum_{k=-n}^n e^{2\pi i k x} = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)},$$

the Dirichlet kernel, and let  $f^N$  be the de la Vallée-Poussin kernel which is defined by  $f^N = 2K_{2N+1} - K_N$  via the Fejér kernel  $K_N = (N+1)^{-1} \sum_{j=0}^N D_j$ . Then  $[\widehat{f^N}]_k = 1$  for  $|k| \leq N+1$  and thus  $S_n f^N = D_n$  for  $|n| \leq N+1$ . We have  $\|f^N\|_{L^1(\mathbb{T})} = O(1)$ , and  $\|f^N\|_\infty = O(N)$  and therefore  $\|f^N\|_{L^{p,q}(\mathbb{T})} = O(N^{1-1/p})$ .

Let  $N \gg 10^3$  and  $8N^{-1} \leq x \leq 1/8$ . Let  $K = K(x)$  be the largest integer  $< Nx$ . Then for  $0 \leq k < 2K(x)$  there are integers  $n_k(x) \leq N$  so that  $(2n_k(x) + 1)x \in (\frac{1}{4} + k, \frac{3}{4} + k)$ , in particular  $n_k(x) < n_{k+1}(x)$  for  $k < 2K(x) - 1$ . Observe  $\sin((2n_{2j}(x) + 1)\pi x) > \sqrt{2}/2$  and  $\sin((2n_{2j+1}(x) + 1)\pi x) < -\sqrt{2}/2$  for  $0 \leq j \leq K(x) - 1$ . This gives

$$\begin{aligned} \left( \sum_{j=0}^{K(x)-1} |S_{n_{2j+1}(x)} f^N - S_{n_{2j}(x)} f^N|^r \right)^{1/r} &= \left( \sum_{j=0}^{K(x)-1} |D_{n_{2j+1}(x)} - D_{n_{2j}(x)}|^r \right)^{1/r} \\ &\geq \frac{K(x)^{1/r} \sqrt{2}}{\sin(\pi x)} \geq cN^{1/r} x^{1/r-1}, \end{aligned}$$

and this implies for large  $N$

$$\frac{\|\mathcal{V}^r(Sf^N)\|_{L^{p,s}}}{\|f^N\|_{L^{p,1}}} \geq c_{p,s} \begin{cases} N^{\frac{1}{p}-\frac{1}{r'}} & \text{if } p < r', \\ (\log N)^{1/s} & \text{if } p = r'. \end{cases}$$

Thus the  $L^p \rightarrow L^p(V^r)$  boundedness does not hold for  $p < r'$ ; moreover the  $L^{r',1} \rightarrow L^{r',s}(V^r)$  boundedness does not hold for  $s < \infty$ .

### 3. THE MODEL OPERATORS

We shall show in appendix A how to deduce Theorem 1.1 from Theorem 1.2. To start the proof of the main Theorem 1.2, we describe some reductions to model operators involving wave packet decompositions.

First, by interpolation it suffices to prove for  $p \geq r'$  the restricted weak type  $L^{p,1} \rightarrow L^{p,\infty}(V_r)$  bound. Next, by the monotone convergence theorem it suffices to estimate  $L^{p,\infty}(V_r)$  on finite  $x$ -intervals  $[-A, A]$ , with constant independent of  $A$ . By another application of the monotone convergence theorem it suffices, for any fixed  $K$ , to prove the  $L^{p,1} \rightarrow L^{p,\infty}([-A, A])$  bound for

$$(5) \quad \sup_{\xi_0 \leq \dots \leq \xi_K} \left( \sum_{\ell=1}^K |\mathcal{S}f(\xi_\ell, x) - \mathcal{S}f(\xi_{\ell-1}, x)|^r \right)^{1/r}$$

where the sup is taken over all  $(\xi_0, \dots, \xi_K)$  with  $\xi_{\ell-1} \leq \xi_\ell$  for  $\ell = 1, \dots, K$ . Moreover, by the density of Schwartz functions in  $L^{p,1}$  it suffices to prove a uniform estimate for all Schwartz functions. Note that for any Schwartz function  $f$  the expression  $\mathcal{S}[f](\xi, x)$  depends continuously on  $(\xi, x)$ . Therefore it suffices to bound the expression analogous to (5) where we impose the strict inequality  $\xi_{\ell-1} < \xi_\ell$  for  $\ell = 1, \dots, K$ . Moreover, by the continuity it suffices for each finite set  $\Xi \subset \mathbb{R}$  to prove bounds for this expression under

the assumption that the  $\xi_\ell$  belong to  $\Xi$ , and we may also assume that  $\Xi$  does not contain any numbers of the form  $n2^m$  with  $m, n \in \mathbb{Z}$  (*i.e.* no endpoints of dyadic intervals).

We may now linearize the variation norm. Fix  $K \in \mathbb{N}$ , measurable real valued functions  $\xi_0(x) < \dots < \xi_K(x)$ , with values in  $\Xi$  and measurable complex valued functions  $a_1(x), \dots, a_K(x)$  satisfying

$$|a_1(x)|^{r'} + \dots + |a_K(x)|^{r'} = 1.$$

Let

$$\mathcal{S}'[f](x) = \sum_{k=1}^K (\mathcal{S}[f](\xi_k(x), x) - \mathcal{S}[f](\xi_{k-1}(x), x)) a_k(x).$$

Theorem 1.2 will now follow from the estimate

$$(6) \quad \|\mathcal{S}'[f]\|_{L^{p,\infty}(\mathbb{R})} \leq C \|f\|_{L^{p,1}(\mathbb{R})}$$

where  $C$  is independent of  $K$ ,  $\Xi$  and the linearizing functions, and where  $f$  is any Schwartz function. Finally, in order to prove (6) for any fixed  $\Xi$  we may assume that  $\widehat{f}$  has compact support in  $\mathbb{R} \setminus \Xi$  since the space of Schwartz functions with this property is dense in  $L^{p,1}$ ,  $1 < p < \infty$ .

Let  $\mathcal{D} = \{[2^k m, 2^k(m+1)) : m, k \in \mathbb{Z}\}$  be the set of dyadic intervals. A *tile* will be any rectangle  $I \times \omega$  where  $I, \omega$  are dyadic intervals, and  $|I||\omega| = 1/2$ . We will write  $\mathcal{S}'$  as the sum of wave packets adapted to tiles, and then decompose the operator into a finite sum of model operators by sorting the wave packets into a finite number of classes. For each  $k$ ,

$$\mathcal{S}[f](\xi_k, x) - \mathcal{S}[f](\xi_{k-1}, x) = \int \mathbb{1}_{(\xi_{k-1}, \xi_k)}(\xi) \widehat{f}(\xi) e^{2\pi i \xi x} d\xi.$$

To suitably express the difference above as a sum of wave packets, we will first need to construct a partition of  $\mathbb{1}_{(\xi_{k-1}, \xi_k)}$  adapted to certain dyadic intervals. The fact that  $(\xi_{k-1}, \xi_k)$  has two boundary points instead of the one from  $(-\infty, \xi_k)$  will necessitate a slightly more involved discretization argument than that in [20].

For any  $\xi < \xi'$ , let  $\mathbf{J}_{\xi, \xi'}$  be the set of maximal dyadic intervals  $J$  such that  $J \subset (\xi, \xi')$  and  $\text{dist}(J, \xi) \geq |J|$ ,  $\text{dist}(J, \xi') \geq |J|$ . Let  $\nu$  be a  $C^\infty$  function from  $\mathbb{R}$  to  $[0, 1]$  which vanishes on  $(-\infty, -10^{-2}]$ , is identically equal to 1 on  $[10^{-2}, \infty)$ , and so that  $\nu'(x) \geq 0$  for  $-10^{-2} < x < 10^{-2}$ . Given an interval  $J = [a, b)$ , and  $i \in \{-1, 0, 1\}$ , define

$$\varphi_{J,i}(\xi) = \nu\left(\frac{\xi - a}{2^i(b-a)}\right) - \nu\left(\frac{\xi - b}{b-a}\right).$$

Thus if  $c(J) = \frac{a+b}{2}$ , the center of  $J$ , then

$$(7) \quad \varphi_{J,i}(\xi) = \nu_i\left(\frac{\xi - c(J)}{|J|}\right) \text{ where } \nu_i(\eta) = \nu(2^{-i}(\eta + \tfrac{1}{2})) - \nu(\eta - \tfrac{1}{2}),$$

and we notice for  $i \in \{-1, 0, 1\}$  both  $\nu_i$  and  $\sqrt{\nu_i}$  are  $C^\infty$  functions supported in  $[-\frac{13}{25}, \frac{13}{25}]$  (more precisely in  $[-\frac{13}{25}, \frac{51}{100}]$ ). Hence  $\varphi_{J,i}$  is supported on a  $\frac{26}{25}$ -dilate of  $J$  with respect to its center.

For each  $J \in \mathbf{J}_{\xi, \xi'}$ , one may check that there is a unique interval  $J' \in \mathbf{J}_{\xi, \xi'}$  which lies strictly to the left of  $J$  and satisfies  $\text{dist}(J', J) = 0$ , and one may check that  $J'$  has size

$|J|/2, |J|$ , or  $2|J|$ . We define  $\varphi_J = \varphi_{J, i(J)}$  where  $i(J)$  is chosen so that  $|J'| = 2^{i(J)}|J|$ . Then

$$(8) \quad \mathbb{1}_{(\xi, \xi')}(\eta) \widehat{f}(\eta) = \sum_{J \in \mathbf{J}_{\xi, \xi'}} \varphi_J(\eta) \widehat{f}(\eta).$$

Since we assume that  $\widehat{f}$  is compactly supported in  $\mathbb{R} \setminus \Xi$  we see that for every pair  $\xi < \xi'$  with  $\xi, \xi' \in \Xi$  only a finite number of dyadic intervals  $J \in \mathbf{J}_{\xi, \xi'}$  are relevant in (8).

We now write each multiplier  $\varphi_J$  as the sum of wave packets. For every tile  $P = I \times J$ , define

$$\phi_P(x) = |I|^{1/2} \mathcal{F}^{-1}[\sqrt{\varphi_J}](x - c(I))$$

where  $c(I)$  is the center of  $I$  and  $\mathcal{F}^{-1}$  denotes the inverse Fourier transform. For each  $J$ , we then have

$$(9) \quad \sum_{|I|=1/(2|J|)} \langle f, \phi_{I \times J} \rangle \widehat{\phi_{I \times J}} = \widehat{f} \varphi_J.$$

To see this we use a Fourier series expansion (cf. [34]). We first observe that  $\widehat{\phi_P}(\xi) = \sqrt{|I|} \sqrt{\varphi_J(\xi)} e^{-2\pi i c(I)\xi}$  and use  $\langle f, \phi_P \rangle = \langle \widehat{f}, \widehat{\phi_P} \rangle$ . Now let parametrize the centers of the dyadic intervals  $I$  of length  $L$  by  $-(k - \frac{1}{2})L$ ,  $k \in \mathbb{Z}$ . Set  $g_J(\omega) := [\sqrt{\varphi_J} \widehat{f}](c(J) + L^{-1}\omega) e^{\pi i \omega}$  and note that  $g_J$  is supported in  $[-\frac{13}{50}, \frac{13}{50}]$ . The left hand side of (9) is equal to

$$\begin{aligned} & \sqrt{\varphi_J(\xi)} \sum_k \int \widehat{f}(\eta) \sqrt{\varphi_J(\eta)} e^{-2\pi i (kL - \frac{L}{2})\eta} L d\eta e^{2\pi i (kL - \frac{L}{2})\xi} \\ &= \sqrt{\varphi_J(\xi)} e^{-\pi i L(\xi - c(J))} \sum_{k \in \mathbb{Z}} \int_{-1/2}^{1/2} g_J(\omega) e^{-2\pi i k \omega} d\omega e^{2\pi i k L(\xi - c(J))} \\ &= \sqrt{\varphi_J(\xi)} e^{-\pi i L(\xi - c(J))} g_J(L(\xi - c(J))) = \widehat{f}(\xi) \varphi_J(\xi) \end{aligned}$$

which gives (9). This in turn yields the representation of  $\mathcal{S}'[f]$  in terms of wave packets:

$$(10) \quad \mathcal{S}'[f](x) = \sum_{k=1}^K \left( \sum_{J \in \mathbf{J}_{\xi_{k-1}(x), \xi_k(x)}} \sum_{|I|=1/(2|J|)} \langle f, \phi_{I \times J} \rangle \phi_{I \times J} \right) a_k(x).$$

For the function  $f$  under consideration the above Fourier series expansion converges in  $L^2$ -Sobolev spaces of arbitrary high order and thus the convergence in (10) is uniform for  $x \in [-A, A]$ . Therefore it suffices, for any *finite* family  $\mathfrak{P}$  of tiles, to consider the operator  $\mathcal{S}''$  defined by

$$(11) \quad \mathcal{S}''[f](x) = \sum_{k=1}^K \left( \sum_{\substack{(I, J) \in \mathfrak{P} \\ J \in \mathbf{J}_{\xi_{k-1}(x), \xi_k(x)}}} \langle f, \phi_{I \times J} \rangle \phi_{I \times J}(x) \right) a_k(x).$$

The wave packets will be sorted into a finite number of classes, each well suited for further analysis. Sorting is accomplished by dividing every  $\mathbf{J}_{\xi, \xi'}$  into a finite number of disjoint sets. These sets will be indexed by a fixed subset of  $\{1, 2, 3\} \times \{1, 2, 3, 4\}^2 \times \{\text{left}, \text{right}\}$ . Specifically, for each  $(m, n, \text{side}) \in \{1, 2, 3, 4\}^2 \times \{\text{left}, \text{right}\}$ , we define

- $\mathbf{J}_{\xi, \xi', (1, m, n, \text{side})} = \{J \in \mathcal{D} : J \subset (\xi, \xi'), \xi \text{ is in the interval } J - (m+1)|J|, \xi' \text{ is in the interval } J + (n+1)|J|, \text{ and } J \text{ is the } \text{side}\text{-child of its dyadic parent}\}.$

- $\mathbf{J}_{\xi, \xi', (2, m, n, \text{side})} = \{J \in \mathcal{D} : J \subset (\xi, \xi'), \xi \text{ is in the interval } J - (m+1)|J|, \text{dist}(\xi', J) \geq n|J|, \text{ and } J \text{ is the } \textit{side}\text{-child of its dyadic parent}\}.$
- $\mathbf{J}_{\xi, \xi', (3, m, n, \text{side})} = \{J \in \mathcal{D} : J \subset (\xi, \xi'), \text{dist}(\xi, J) > m|J|, \xi' \text{ is in the interval } J + (n+1)|J|, \text{ and } J \text{ is the } \textit{side}\text{-child of its dyadic parent}\}.$

We will choose  $R \subset \{1, 2, 3\} \times \{1, 2, 3, 4\}^2 \times \{\text{left}, \text{right}\}$  so that for each  $\xi, \xi'$ , the collection  $\{\mathbf{J}_{\xi, \xi', \rho}\}_{\rho \in R}$  is pairwise disjoint and  $\mathbf{J}_{\xi, \xi'} = \cup_{\rho \in R} \mathbf{J}_{\xi, \xi', \rho}$ . We will also assume that for each  $\rho \in R$  there is an  $i(\rho) \in \{-1, 0, 1\}$  such that  $|J'| = 2^{i(\rho)}|J|$  for every  $\xi < \xi'$ ,  $J \in \mathbf{J}_{\xi, \xi', \rho}$  and  $J' \in \mathbf{J}_{\xi, \xi'}$  with  $J'$  strictly to the left of  $J$  and  $\text{dist}(J, J') = 0$ . One may check that these conditions are satisfied, say, for

$$R = \{(1, 2, 1, \text{left}), (1, 2, 2, \text{left}), (1, 3, 1, \text{left}), \\ (1, 3, 2, \text{left}), (2, 1, 1, \text{left}), (2, 1, 1, \text{right}), (2, 2, 1, \text{right}), \\ (3, 4, 1, \text{left}), (3, 3, 1, \text{right}), (3, 4, 2, \text{left})\}.$$

It now follows that

$$\mathcal{S}''[f] = \sum_{\rho \in R} \mathcal{S}^\rho[f]$$

where

$$\mathcal{S}^\rho[f](x) = \sum_{k=1}^K \left( \sum_{\substack{(I \times J) \in \mathfrak{P} \\ J \in \mathbf{J}_{\xi_{k-1}(x), \xi_k(x), \rho}}} \langle f, \phi_{I \times J} \rangle \phi_{I \times J} \right) a_k(x).$$

It will be convenient to rewrite each operator  $\mathcal{S}^\rho$  in terms of *multitiles*. A multitile will be a subset of  $\mathbb{R}^2$  of the form  $I \times \omega$  where  $I \in \mathcal{D}$  and where  $\omega$  is the union of three intervals  $\omega_l, \omega_u, \omega_h$  in  $\mathcal{D}$ . For each  $\rho = (i, m, n, \text{side}) \in R$ , we consider a set of  $\rho$ -multitiles which is parameterized by  $\{(I, \omega_u) : I, \omega_u \in \mathcal{D}, |I||\omega_u| = 1/2, \text{ and } \omega_u \text{ is the } \textit{side}\text{-child of its parent}\}$ . Specifically, given  $\omega_u = [a, b]$

- If  $\rho = (1, m, n, \text{side})$  then  $\omega_l = \omega_u - (m+1)|\omega_u|$  and  $\omega_h = \omega_u + (n+1)|\omega_u|$ .
- If  $\rho = (2, m, n, \text{side})$  then  $\omega_l = \omega_u - (m+1)|\omega_u|$  and  $\omega_h = [a + (n+1)|\omega_u|, \infty)$ .
- If  $\rho = (3, m, n, \text{side})$  then  $\omega_l = (-\infty, b - (m+1)|\omega_u|)$  and  $\omega_h = \omega_u + (n+1)|\omega_u|$ .

For  $i = 1, 2, 3$  we shall say that  $\rho$  is an  $i$ -index if  $\rho = (i, m, n, \text{side})$ . For every  $\rho$ -multitile  $P$ , let  $a_P(x) = a_k(x)$  if  $k$  satisfies  $1 \leq k \leq K$  and  $\xi_{k-1}(x) \in \omega_l$  and  $\xi_k(x) \in \omega_h$  (such a  $k$  would clearly be unique), and  $a_P(x) = 0$  if there is no such  $k$ . Then,

$$\mathcal{S}^\rho[f](x) = \sum_{P \in \mathbf{P}_\rho} \langle f, \phi_P \rangle \phi_P(x) a_P(x)$$

where, for each  $\rho$ -multitile  $P$ ,  $\phi_P(x) = \sqrt{|I|} \mathcal{F}^{-1}[\sqrt{\varphi_{\omega_u, i(\rho)}}](x - c(I))$  and  $\mathbf{P}_\rho$  denotes the set of all  $\rho$ -multitiles for which  $I \times \omega_u$  belongs to  $\mathfrak{P}$ .

Inequality (6) and hence Theorem 1.2 will then follow after proving the bound

$$(12) \quad \|\mathcal{S}^\rho[f]\|_{L^{p, \infty}} \lesssim \|f\|_{L^{p, 1}}$$

for each  $\rho \in R$ . We shall only give the proof of this estimate for the case that  $\rho$  is a 1-index or  $\rho$  is a 2-index, and the case where  $\rho$  is a 3-index can be deduced by symmetry considerations. Indeed, if  $P = (I, \omega_u) := ([a, b], [c, d])$  and  $\tilde{P} := ([-b, a], [-d, c])$  then

$\langle f, \phi_P \rangle \phi_P = \langle f(-\cdot), \tilde{\phi}_{\tilde{P}} \rangle \tilde{\phi}_{\tilde{P}}$  where, in the definition of the  $\tilde{\phi}_{\tilde{P}}$  the function  $\nu_i$  in (7) is replaced with  $\nu_i(-\cdot)$  (both are supported in  $(-\frac{13}{25}, \frac{13}{25})$ ). Now reflection sends a half open interval  $[a, b)$  to a half open interval  $(-b, -a]$ , however this plays no role in our symmetry argument if, as we do, we assume that the set  $\Xi$  does not contain endpoints of dyadic intervals. We then see that the estimation of  $S^\rho[f]$  for  $\rho = (3, m, n, \text{side})$  is equivalent to the estimation of a  $S^{\tilde{\rho}}[f(-\cdot)]$  where the corresponding set  $\{P\}$  of multitiles is replaced with a set  $\{\tilde{P}\}$  of index  $\tilde{\rho} = (2, n, m, \text{opposite side})$  and the set  $\Xi$  is replaced with  $\{\xi : -\xi \in \Xi\}$ . Beginning with (14) both  $\nu_i$  or  $\nu_i(-\cdot)$  are allowed in the definition of the functions  $\varphi_{J,i}$  and  $\phi_P$ .

By the usual characterization of  $L^{p,1}$  as superpositions of functions bounded by characteristic functions it suffices to show that

$$\text{meas}(\{x : |S^\rho[f](x)| > \lambda\}) \leq C^p \lambda^{-p} |F|$$

where  $F \subset \mathbb{R}$ ,  $|F| > 0$ ,  $|f| \leq \mathbb{1}_F$ ,  $\lambda > 0$ ,  $2 < r < \infty$ , and  $r' \leq p < (1/2 - 1/r)^{-1}$ . This is accomplished by proving that for every measurable  $E \subset \mathbb{R}$ ,

$$(13) \quad \text{meas}\left(\{x \in E : |S^\rho[f](x)| > C(|F|/|E|)^{1/p}\}\right) \leq |E|/2.$$

Indeed, if we set  $E_\lambda = \{x : |S^\rho[f](x)| > \lambda\}$  then by the finiteness of the set of tiles under consideration the set  $E_\lambda$  has a priori finite measure. If  $|E_\lambda| \leq C^p \lambda^{-p} |F|$  then there is nothing to prove. If the opposite inequality  $|E_\lambda| > C^p \lambda^{-p} |F|$  were true then  $\lambda > C(|F|/|E_\lambda|)^{1/p}$  and inequality (13) applied to  $E = E_\lambda$  would yield that  $|E_\lambda| \leq |E_\lambda|/2$ , a contradiction.

We finally note that, after possibly rescaling, we may assume that  $1 \leq |E| \leq 2$  in (13). The next four sections will be devoted to the proof of inequality (13) in this case.

#### 4. ENERGY AND DENSITY

Recall that  $S^\rho[f](x) = \sum_P \langle f, \phi_P \rangle \phi_P a_P$  where  $P$  ranges over an arbitrary finite collection of  $\rho$ -multitiles,  $\rho$  is a 1 or 2-index. It is our goal to show (13) and for this and the next chapter we fix the function  $f$  with  $|f| \leq \mathbb{1}_F$  and the set  $E$ .

Fix  $1 \leq C_3 < C_2 < C_1$ , with  $C_2 \in \mathbb{N}$ , such that for every multitile  $P$ ,

$$\begin{aligned} \text{supp}(\widehat{\phi}_P) &\subset C_3 \omega_u \\ C_2 \omega_u \cap C_2 \omega_l &= \emptyset, \quad C_2 \omega_u \cap \omega_h = \emptyset, \\ C_2 \omega_l &\subset C_1 \omega_u, \quad C_2 \omega_u \subset C_1 \omega_l; \end{aligned}$$

recall that dilations of finite intervals are with respect to their center. One may check that the values  $C_3 = 11/10$ ,  $C_2 = 2$ , and  $C_1 = 12$  satisfy all these properties.

The wave packet is adapted to the multitile  $P$ . As  $\widehat{\phi}_P$  is compactly supported (in  $C_3 \omega_u$ ) the function  $\phi_P$  cannot have compact support, but as a replacement we have the following bounds involving

$$w_I(x) := \frac{1}{|I|} \left(1 + \frac{|x - c(I)|}{|I|}\right)^{-N}$$

for a fixed large  $N \gg 10$ ; namely

$$(14) \quad \left| \frac{d^n}{dx^n} (\exp(-2\pi i c(\omega_u) \cdot) \phi_P)(x) \right| \leq C'(n) |I|^{(1/2)-n} |w_I(x)|$$

for each  $n \geq 0$ .

We are working with a given finite set of  $\rho$ -multitiles  $\mathbf{P}$  (with  $\rho$  a 1- or 2-index) and we let  $M_o = M_o(\mathbf{P})$  be the smallest integer  $M$  for which all tiles are contained in the square  $[-2^M, 2^M]^2$ . Throughout this paper we fix

$$\Xi^{\text{top}} = \{ \eta : |\eta| \leq C_1 2^{M_o+10}, \eta = n 2^{-M_o-10}, \text{ for some } n \in \mathbb{Z} \}$$

as the set of admissible top-frequencies for trees, as in the following definition.

**Definition.** Consider a triple  $\mathcal{T} = (T_{\mathcal{T}}, I_{\mathcal{T}}, \xi_{\mathcal{T}})$ , with a set of multitiles  $T_{\mathcal{T}}$ , a dyadic interval  $I_{\mathcal{T}} \subset [-2^{M_o}, 2^{M_o})$  and a point  $\xi_{\mathcal{T}} \in \Xi^{\text{top}}$ . We say that  $\mathcal{T}$  is a tree if the following properties are satisfied.

(i)  $I \subset I_{\mathcal{T}}$  for all  $P = (I, \omega_u) \in T_{\mathcal{T}}$ .

(ii) If  $P = (I, \omega_u)$  and  $\omega_m$  denotes the convex hull of  $C_2 \omega_u \cup C_2 \omega_l$  then

$$\omega_{\mathcal{T}} := \left[ \xi_{\mathcal{T}} - \frac{C_2-1}{4|I_{\mathcal{T}}|}, \xi_{\mathcal{T}} + \frac{C_2-1}{4|I_{\mathcal{T}}|} \right)$$

is contained in  $\omega_m$ .

We refer to  $I_{\mathcal{T}}$  as the top interval of the tree, and to  $\xi_{\mathcal{T}}$  as the top frequency of the tree.

In order not to overload the notation we usually refer to the set  $T$  as “the” tree (keeping in mind that it carries additional information of a top frequency and a top interval), and we shall also use the notation  $I_T$ ,  $\xi_T$  and  $\omega_T$  in place of  $I_{\mathcal{T}}$ ,  $\xi_{\mathcal{T}}$  and  $\omega_{\mathcal{T}}$ . With this convention we also define

**Definition.**

(i) A tree  $(T, I_T, \xi_T)$  is  $l$ -overlapping if  $\xi_T \in C_2 \omega_l$  for every  $P \in T$ .

(ii) A tree  $(T, I_T, \xi_T)$  is  $l$ -lacunary if  $\xi_T \notin C_2 \omega_l$  for every  $P \in T$ .

Notice that the union of two trees with the top data  $I_T, \xi_T$  is again a tree with the same top data. Also, the union of two  $l$ -overlapping trees with the same top data is again an  $l$ -overlapping tree with the same top data.

We split our finite collection of multitiles into a bounded number of subcollections satisfying certain separation conditions (i.e. henceforth all multitiles will be assumed to belong to a fixed subcollection).

**Separation assumptions.**

$$(15) \quad \text{If } P, P' \text{ satisfy } |\omega'_u| < |\omega_u|, \text{ then } |\omega'_u| \leq \frac{C_2 - C_3}{2C_1} |\omega_u|.$$

$$(16) \quad \text{If } P, P' \text{ satisfy } C_1 \omega_u \cap C_1 \omega'_u \neq \emptyset \text{ and } |\omega_u| = |\omega'_u| \text{ then } \omega_u = \omega'_u.$$

As immediate but important consequence of the separation assumptions is the frequently used

**Observation 4.1.** *Let  $T$  be a tree satisfying the separation properties (15) and (16). Then the following properties hold.*

- (i) *If  $T$  is an  $l$ -overlapping tree,  $P, P' \in T$ , and  $|\omega'_u| < |\omega_u|$  then  $C_3\omega_u \cap C_2\omega'_u = \emptyset$ .*
- (ii) *If  $T$  is an  $l$ -lacunary tree,  $P, P' \in T$ , and  $|\omega'_u| < |\omega_u|$  then  $C_3\omega_l \cap C_3\omega'_l = \emptyset$ .*
- (iii) *If  $P, P' \in T$ ,  $P \neq P'$ , and  $|\omega_u| = |\omega'_u|$ , then  $I \cap I' = \emptyset$ .*

As in previous proofs of Carleson's theorem (in particular [20]) we shall split the set of multitiles into subsets with controllable *energy* and *density* associated to the function  $f$  and the set  $E$ , respectively. Here we work with the following definitions.

**Definition.** Fix  $f \in L^2(\mathbb{R})$  and a measurable set  $E \subset \mathbb{R}$ . Given any collection of multitiles  $\mathbf{P}$  we define

$$\text{energy}(\mathbf{P}) = \sup_T \sqrt{\frac{1}{|I_T|} \sum_{P \in T} |\langle f, \phi_P \rangle|^2}$$

where the sup ranges over all  $l$ -overlapping trees  $T \subset \mathbf{P}$ .

Given a measurable set  $E \subset \mathbb{R}$  we set

$$\text{density}(\mathbf{P}) = \sup_T \left( \frac{1}{|I_T|} \int_E \left( 1 + \frac{|x - c(I_T)|}{|I_T|} \right)^{-4} \sum_{k=1}^K |a_k(x)|^{r'} \mathbb{1}_{\omega_T}(\xi_{k-1}(x)) \, dx \right)^{1/r'}$$

where the sup is over all non-empty trees  $T \subset \mathbf{P}$ .

*Remark.* Concerning the terminology, one can argue that the squareroot should be omitted in the definition of an energy. However we work with the above definition to conform to [20] and other papers in time-frequency analysis.

**Lemma 4.2.** *Let  $|f|$  be bounded by  $\mathbb{1}_F$ . For any family  $\mathbf{P}$  of multitiles the density of  $\mathbf{P}$  (with respect to the set  $E$ ) and the energy (with respect to  $f$ ) are bounded by a universal constant.*

*Proof.* Clearly the density is bounded by  $\int_{\mathbb{R}} (1 + |x|)^{-4} dx < 3$ . Concerning the energy bound we let  $T$  be any  $l$ -overlapping tree, and split  $f = f' + f''$  where  $f' = \mathbb{1}_{3I_T} f$ . We estimate  $\sum_{P \in T} |\langle f', \phi_P \rangle|^2 \leq \|f'\|_{L^2}^2 (\sum_{P \in T} |\langle f', \phi_P \rangle|^2)^{1/2}$ , by the Cauchy-Schwarz inequality. Now use that the supports of the  $\widehat{\phi}_P$  are disjoint for different sizes of frequency intervals, and then, for a fixed size, use the bounds (14) (for  $n = 0$ ) to see that

$$\left\| \sum_{P \in T} \langle f', \phi_P \rangle \phi_P \right\|_{L^2} \leq \left( \sum_{j \in \mathbb{Z}} \left\| \sum_{\substack{P \in T \\ |\omega_u| = 2^{-j}}} \langle f', \phi_P \rangle \phi_P \right\|_{L^2}^2 \right)^{1/2} \lesssim \left( \sum_{P \in T} |\langle f', \phi_P \rangle|^2 \right)^{1/2}.$$

Hence,

$$\sum_{P \in T} |\langle f', \phi_P \rangle|^2 \lesssim \|f'\|_{L^2}^2 \leq |F \cap 3I_T| \lesssim |I_T|.$$

Furthermore, since  $|f''| \leq \mathbb{1}_{\mathbb{R} \setminus 3I_T}$ , we have the estimate

$$|\langle f'', \phi_P \rangle| \lesssim |I|^{1/2} (1 + \text{dist}(I, \mathbb{R} \setminus 3I_T)/|I|)^{-(N-1)} \lesssim |I|^{1/2} (|I|/|I_T|)^{N-1}$$

Summing in  $P$ , we obtain

$$\sum_{P \in T} |\langle f'', \phi_P \rangle|^2 \lesssim |I_T|.$$

Combining the estimates for  $f'$  and  $f''$  we see that  $\sum_{P \in T} |\langle f, \phi_P \rangle|^2 \lesssim |I_T|$  for every  $l$ -overlapping tree and it follows that the energy of  $\mathbf{P}$  with respect to  $f$  is bounded above by a universal constant.  $\square$

The following proposition allows one to decompose an arbitrary collection of multitiles into the union of trees, where the trees are divided into collections  $\mathbf{T}_j$  with the energy of trees from  $\mathbf{T}_j$  bounded by  $2^{-j}$ . The control over energy is balanced by an  $L^q$  bound for the functions  $N_{j,\ell} := \sum_{T \in \mathbf{T}_j} \mathbb{1}_{2^\ell I_T}$ . In contrast to [20] and [12], it is necessary here to consider  $q > 1$  and  $\ell > 0$  in order to effectively use the tree estimate Proposition 5.1 with  $q > 1$ . Note that such  $L^q$  bound for  $N_{j,\ell}$  is established by combining (17) and (18) below. The bound (20) permits one to make further decompositions to take advantage of large  $|F|$  in the  $L^q$  bound for the  $N_{j,\ell}$  while maintaining compatibility with bounds for trees with a fixed density obtained from Proposition 4.4.

**Proposition 4.3.** *Let  $\mathcal{E} > 0$ , let  $|f|$  be bounded above by  $\mathbb{1}_F$  and let  $\mathbf{P}$  be a collection of multitiles with energy bounded above by  $\mathcal{E}$ . Then, there is a collection of trees  $\mathbf{T}$  such that*

$$(17) \quad \sum_{T \in \mathbf{T}} |I_T| \lesssim \mathcal{E}^{-2} |F|$$

and

$$\text{energy} \left( \mathbf{P} \setminus \bigcup_{T \in \mathbf{T}} T \right) \leq \mathcal{E}/2,$$

and such that, for every integer  $\ell \geq 0$ ,

$$(18) \quad \left\| \sum_{T \in \mathbf{T}} \mathbb{1}_{2^\ell I_T} \right\|_{BMO} \lesssim 2^{2\ell} \mathcal{E}^{-2}.$$

Furthermore, if for some collection of trees  $\mathbf{T}'$ ,

$$(19) \quad \mathbf{P} = \bigcup_{T' \in \mathbf{T}'} T'$$

then

$$(20) \quad \sum_{T \in \mathbf{T}} |I_T| \lesssim \sum_{T' \in \mathbf{T}'} |I_{T'}|.$$

Above, and subsequently,  $\|\cdot\|_{BMO}$  denotes the dyadic  $BMO$  norm.

*Proof.* We select trees through an iterative procedure. First, if  $\text{energy}(\mathbf{P}) \leq \mathcal{E}/2$  then no tree is chosen and  $\mathbf{T} = \emptyset$ .



If  $\text{energy}(\mathbf{P}) \geq \mathcal{E}/2$  then we observe that there is an  $l$ -overlapping tree  $S \subset \mathbf{P}$  for which

$$(21) \quad \frac{1}{|I_S|} \sum_{P \in S} |\langle f, \phi_P \rangle|^2 \geq \mathcal{E}^2/4.$$

There is only a finite number of such trees and as  $S_1$  we choose one for which the top datum  $\xi_S$  is maximal (in  $\mathbb{R}$ ). Note that the maximality can be achieved as we restrict all top frequency data to the finite set  $\Xi^{\text{top}}$ . Let  $T_1$  be the tree in  $\mathbf{P}$  which has top data  $(\xi_{S_1}, I_{S_1})$  and which is maximal with respect to inclusion.

Suppose that trees  $S_k, T_k$  have been chosen for  $k = 1, \dots, j$ . Set

$$\mathbf{P}_j = \mathbf{P} \setminus \bigcup_{k=1}^j T_k$$

If  $\text{energy}(\mathbf{P}_j) \leq \mathcal{E}/2$  then we terminate the procedure, set  $\mathbf{T} = \{T_k\}_{1 \leq k \leq j}$  and  $n = j$ . Otherwise, we may find an  $l$ -overlapping tree  $S \subset \mathbf{P}_j$  such that (21) holds. Among  $l$ -overlapping in  $\mathbf{P}_j$  satisfying (21) choose one with maximal top-frequency (in  $\Xi^{\text{top}}$ ) and label this tree  $S_{j+1}$ . Let  $T_{j+1}$  be the maximal tree in  $\mathbf{P}_j$  which has top data  $(\xi_{S_{j+1}}, I_{S_{j+1}})$  and which is maximal with respect to inclusion. This process will eventually stop since each  $T_j$  is nonempty and  $\mathbf{P}$  is finite.

*Proof of (17).* It suffices to show

$$(22) \quad \left( \frac{\mathcal{E}^2}{|F|} \sum_{j=1}^n |I_{S_j}| \right)^2 \lesssim \frac{\mathcal{E}^2}{|F|} \sum_{j=1}^n |I_{S_j}|.$$

Since the  $S_j$  satisfy (21), we have

$$\left( \frac{\mathcal{E}^2}{|F|} \sum_{j=1}^n |I_{S_j}| \right)^2 \leq 16|F|^{-2} \left( \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \right)^2.$$

Now

$$\begin{aligned} \left( \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \right)^2 &= \left( \left\langle \sum_{j=1}^n \sum_{P \in S_j} \langle f, \phi_P \rangle \phi_P, f \right\rangle \right)^2 \\ &\leq \|f\|_2^2 \left\| \sum_{j=1}^n \sum_{P \in S_j} \langle f, \phi_P \rangle \phi_P \right\|_2^2 \leq |F| \sum_{j=1}^n \sum_{P \in S_j} \sum_{k=1}^n \sum_{P' \in S_k} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \end{aligned}$$

where in the last inequality we used  $|f| \leq \mathbb{1}_F$ . By symmetry, it remains, for (22) to show that

$$(23) \quad \sum_{j=1}^n \sum_{k=1}^n \sum_{P \in S_j} \sum_{P' \in S_k: |I'|=|I|} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \lesssim \mathcal{E}^2 \sum_{j=1}^n |I_{S_j}|$$

and

$$(24) \quad \sum_{j=1}^n \sum_{k=1}^n \sum_{P \in S_j} \sum_{P' \in S_k: |I'| < |I|} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \lesssim \mathcal{E}^2 \sum_{j=1}^n |I_{S_j}|.$$

In both cases, we will use the estimate

$$(25) \quad |\langle \phi_P, \phi_{P'} \rangle| \lesssim \left( \frac{|I|}{|I'|} \right)^{1/2} \langle w_I, \mathbb{1}_{I'} \rangle.$$

which holds whenever  $|I'| \leq |I|$ .

Estimating the product of two terms by the square of their maximum, we see that the left side of (23) is

$$\leq 2 \sum_{j=1}^n \sum_{k=1}^n \sum_{P \in S_j} \sum_{P' \in S_k: |I'|=|I|} |\langle f, \phi_P \rangle|^2 |\langle \phi_P, \phi_{P'} \rangle|.$$

Recall that  $\langle \phi_P, \phi_{P'} \rangle = 0$  unless  $C_3\omega_u \cap C_3\omega'_u \neq \emptyset$ . Thus, by (16), (25) and the fact that the  $S_k$  are pairwise disjoint, we can estimate the last display by

$$2 \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \sum_{I': |I'|=|I|} \langle w_I, \mathbb{1}_{I'} \rangle \lesssim \sum_{j=1}^n \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \lesssim \sum_{j=1}^n \mathcal{E}^2 |I_{S_j}|.$$

This finishes the proof of (23).

Applying the Cauchy-Schwarz inequality, we see that the left side of (24) is bounded by

$$\sum_{j=1}^n \left( \sum_{P \in S_j} |\langle f, \phi_P \rangle|^2 \right)^{1/2} \left( \sum_{P \in S_j} \left( \sum_{k=1}^n \sum_{P' \in S_k: |I'| < |I|} |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \right)^2 \right)^{1/2}.$$

Twice using the fact that the energy of  $\mathbf{P}$  is bounded by  $\mathcal{E}$ , we see that the last display is

$$\leq \mathcal{E}^2 \sum_{j=1}^n |I_{S_j}|^{1/2} \left( \sum_{P \in S_j} \left( \sum_{k=1}^n \sum_{P' \in S_k: |I'| < |I|} |\langle \phi_P, |I_{P'}|^{1/2} \phi_{P'} \rangle| \right)^2 \right)^{1/2}.$$

Thus, to prove (24) it remains to show that, for each  $j$ ,

$$(26) \quad \sum_{P \in S_j} \left( \sum_{k=1}^n \sum_{P' \in S_k: |I'| < |I|} |\langle \phi_P, |I_{P'}|^{1/2} \phi_{P'} \rangle| \right)^2 \lesssim |I_{S_j}|.$$

Again, we only have  $|\langle \phi_P, |I_{P'}|^{1/2} \phi_{P'} \rangle|$  nonzero when  $C_3\omega_u \cap C_3\omega'_u \neq \emptyset$  which can only happen if  $\sup C_3\omega_u \in C_3\omega'_u$  or  $\inf C_3\omega_u \in C_3\omega'_u$ . Applying (25), we thus see that the left side of (26) is dominated by a constant times the expression

$$\sum_{P \in S_j} |I_P| \left( \sum_{k=1}^n \sum_{\substack{P' \in S_k: |I'| < |I| \\ \sup C_3\omega_u \in C_3\omega'_u}} \langle w_I, \mathbb{1}_{I'} \rangle \right)^2 + \sum_{P \in S_j} |I_P| \left( \sum_{k=1}^n \sum_{\substack{P' \in S_k: |I'| < |I| \\ \inf C_3\omega_u \in C_3\omega'_u}} \langle w_I, \mathbb{1}_{I'} \rangle \right)^2.$$

We now claim that, for each  $P \in S_j$  (with time interval  $I$ ),

$$(27) \quad \left( \sum_{k=1}^n \sum_{\substack{P' \in S_k: |I'| < |I| \\ \sup C_3\omega_u \in C_3\omega'_u}} \langle w_I, \mathbb{1}_{I'} \rangle \right)^2 \leq \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle$$

and that the same inequality with sup replaced by inf in the  $P'$  summation holds as well. To see this consider two multitiles  $P^1 = (I^1, \omega_u^1) \in S_{\kappa_1}$  and  $P^2 = (I^2, \omega_u^2) \in S_{\kappa_2}$ ,  $P^1 \neq P^2$ . so that  $|I^2| \leq |I^1|$  and  $C_3\omega_u^1 \cap C_3\omega_u^2 \neq \emptyset$ . The last condition implies  $\kappa_1 \neq \kappa_2$  (since  $S_{\kappa_1}$  and  $S_{\kappa_2}$  are  $l$ -overlapping). The inequality (27) is immediate if we can show that  $I^1$  and  $I^2$  are disjoint and if in addition  $|I^2| < |I^1|$ , then  $I_2$  does not belong to the top interval of the tree  $S_{\kappa_1}$ . Now, if  $|I^1| = |I^2|$ , then from (16) it follows that  $\omega_u^1 = \omega_u^2$  and hence, since  $P^1 \neq P^2$ , we have  $I^1 \cap I^2 = \emptyset$ . If  $|I^2| < |I^1|$ , then by (15)  $|\omega_u^1| \leq \frac{C_2 - C_3}{2C_1} |\omega_u^2|$ ; since  $C_3\omega_u^1 \cap C_3\omega_u^2 \neq \emptyset$  this implies that  $\inf C_2\omega_l^1 > \sup C_2\omega_l^2$ . As both trees are  $l$ -overlapping the top frequency of  $S_{\kappa_1}$  belongs to  $C_2\omega_l^1$  and is above the

top frequency of  $S_{\kappa_2}$  which belongs to  $C_2\omega_l^2$ . Thus by the maximality condition on the top frequency in the selection process of the trees we see that the tree  $S_{\kappa_1}$  was selected before the tree  $S_{\kappa_2}$ , i.e.  $\kappa_1 < \kappa_2$ . This implies that  $P_2$  does not belong to the tree  $T_{\kappa_1}$ , and since the interval  $\omega_{S_{\kappa_1}}$  is contained in the convex hull of  $\omega_l^2$  and  $\omega_u^2$  we see that the time intervals  $I^2$  and  $I_{T_{\kappa_1}} = I_{S_{\kappa_1}}$  cannot intersect. Thus  $I^1 \cap I^2 = \emptyset$  as  $I^1 \subset I_{T_{\kappa_1}}$ . This concludes the argument for (27).

Now by the disjointness condition we see that indeed the left hand side of (26) is bounded by a constant times

$$\begin{aligned} \sum_{P \in S_j} |I_P| \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle^2 &\lesssim \sum_{\ell: 2^\ell \leq |I_{S_j}|} 2^\ell \sum_{P \in S_j: |I|=2^\ell} \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle \\ &\lesssim |I_{S_j}| \sup_{\ell: 2^\ell \leq |I_{S_j}|} \sum_{P \in S_j: |I|=2^\ell} \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle \end{aligned}$$

One may check that, for each  $\ell$ ,

$$\sum_{P \in S_j: |I|=2^\ell} \langle w_I, \mathbb{1}_{\mathbb{R} \setminus I_{S_j}} \rangle \lesssim 1$$

and (26) follows. We have already seen that (26) implies (24); this completes the proof of (17).

*Proof of (18).* We need to show that for each dyadic interval  $J$ , we have

$$\frac{1}{|J|} \int_J \left| \sum_{T \in \mathbf{T}} \mathbb{1}_{2^\ell I_T}(x) - \frac{1}{|J|} \int_J \sum_{T \in \mathbf{T}} \mathbb{1}_{2^\ell I_T}(y) dy \right| dx \lesssim 2^{2\ell} \mathcal{E}^{-2}.$$

This is an immediate consequence of

$$(28) \quad \sum_{\substack{T \in \mathbf{T} \\ J \cap 2^\ell I_T \neq \emptyset, J}} |I_T| \lesssim \mathcal{E}^{-2} 2^\ell |J|.$$

Let

$$\tilde{\mathbf{T}} = \{T \in \mathbf{T} : I_T \subset 2^{\ell+1}J, |I_T| \leq |J|\}$$

and note that if  $T \in \mathbf{T}$  with  $2^\ell I_T \cap J \neq \emptyset, J$  then  $T \in \tilde{\mathbf{T}}$ . Write  $f = f' + f''$  where  $|f'| \leq \mathbb{1}_{F \cap 2^{\ell+5}J}$  and  $|f''| \leq \mathbb{1}_{F \cap \mathbb{R} \setminus 2^{\ell+5}J}$ .

We will write  $\tilde{\mathbf{T}}$  as the union of collections of trees  $\mathbf{T}^{\text{main}} \cup \mathbf{T}^0 \cup \mathbf{T}^1 \cup \dots$  each of which will have certain properties related to the energy. For each tree  $T \in \tilde{\mathbf{T}}$  there is an  $l$ -overlapping tree  $S = S(T)$  chosen in the algorithm above with  $I_S = I_T$  and

$$(29) \quad \frac{1}{|I_S|} \sum_{P \in S} |\langle f, \phi_P \rangle|^2 \geq \mathcal{E}^2/4.$$

Let

$$\mathbf{T}^0 = \{T \in \tilde{\mathbf{T}} : \frac{1}{|I_S|} \sum_{P \in S} |\langle f'', \phi_P \rangle|^2 \geq \mathcal{E}^2/16\}.$$

For  $j \geq 1$ , define

$$\mathbf{T}^j = \{T \in \tilde{\mathbf{T}} : \sup_{\substack{S' \subset S(T) \\ |I_{S'}|=2^{-j}|I_T|}} \frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f'', \phi_P \rangle|^2 \geq \mathcal{E}^2/16\}$$

where, for each  $T$ , the sup above is taken over all  $l$ -overlapping trees  $S'$  with  $S' \subset S(T)$ . Finally, let

$$\mathbf{T}^{\text{main}} = \{T \in \tilde{\mathbf{T}} \setminus (\mathbf{T}^0 \cup \mathbf{T}^1 \cup \dots)\}.$$

We split the sum (28) into the “main” term involving trees in  $\mathbf{T}^{\text{main}}$  and an error term involving  $\cup_{j \geq 0} \mathbf{T}_j$ .

We first consider the main term. It is our objective to prove

$$(30) \quad \sum_{T \in \mathbf{T}^{\text{main}}} |I_T| \lesssim \mathcal{E}^{-2} 2^\ell |J|.$$

Let  $T \in \mathbf{T}^{\text{main}}$  and let  $S'$  be any  $l$ -overlapping tree contained in  $S$  satisfying  $|I_{S'}| \leq |I_S|$ . Since the energy of  $\mathbf{P}$  is bounded by  $\mathcal{E}$  and since  $T$  is not in any  $\mathbf{T}^j$ , we have

$$\frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f', \phi_P \rangle|^2 \leq 2 \frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f, \phi_P \rangle|^2 + 2 \frac{1}{|I_{S'}|} \sum_{P \in S'} |\langle f'', \phi_P \rangle|^2 \lesssim \mathcal{E}^2.$$

From (29) and the fact that  $T \notin \mathbf{T}^0$ , we have

$$\frac{1}{|S(T)|} \sum_{P \in S(T)} |\langle f', \phi_P \rangle|^2 \geq \mathcal{E}^2/8 - \mathcal{E}^2/16 = \mathcal{E}^2/16.$$

This inequality allows us to essentially repeat the above proof of (17). The  $l$ -overlapping trees  $S(T)$  form a (finite or infinite) subsequence of the sequence  $S_j$  which we denote by  $S_{j(\nu)}$  so that we have  $j(\nu) > j(\nu')$  for  $\nu > \nu'$ . We need to prove the analogue of (22) which is

$$(31) \quad \frac{\mathcal{E}^2}{|F \cap 2^{\ell+5} J|} \sum_{\nu=1}^n |I_{S_{j(\nu)}}| \lesssim 1$$

and as before we are aiming to estimate the square of the expression on the left hand side by the expression itself. The  $S_{j(\nu)}$  satisfy

$$\frac{1}{|S_{j(\nu)}|} \sum_{P \in S_{j(\nu)}} |\langle f', \phi_P \rangle|^2 \geq \frac{\mathcal{E}^2}{16}$$

and therefore

$$\left( \frac{\mathcal{E}^2}{|F \cap 2^{\ell+5} J|} \sum_{\nu=1}^n |I_{S_{j(\nu)}}| \right)^2 \leq 256 |F \cap 2^{\ell+5} J|^{-2} \left( \sum_{\nu=1}^n \sum_{P \in S_{j(\nu)}} |\langle f, \phi_P \rangle|^2 \right)^2.$$

We continue to argue with exactly the same reasoning as in the proof of (22), replacing  $f$  with  $f'$  and  $F$  with  $F \cap 2^{\ell+5} J$ . This leads to the proof of (31) and thus to

$$\sum_{T \in \mathbf{T}^{\text{main}}} |I_T| \lesssim \mathcal{E}^{-2} |F \cap 2^{\ell+5} J|$$

which is clearly  $\lesssim \mathcal{E}^{-2} 2^\ell |J|$ . Thus (30) is established.

For the complimentary terms we prove better estimates, namely, for  $j = 0, 1, 2, \dots$ ,

$$(32) \quad \sum_{T \in \mathbf{T}_j} |I_T| \lesssim 2^{-\ell-j} \mathcal{E}^{-2} |J|.$$

For each  $T \in \mathbf{T}^j$  and  $P \in S(T)$  we have  $|I| \leq 2^{-j}|I_T| \leq 2^{-j}|J|$ . Thus

$$\sum_{T \in \mathbf{T}^j} |I_T| \lesssim \sum_{T \in \mathbf{T}^j} 2^j \mathcal{E}^{-2} \sum_{P \in S(T): |I| \leq 2^{-j}|J|} |\langle f'', \phi_P \rangle|^2.$$

Since the  $S$  are pairwise disjoint, the right hand side is

$$\lesssim 2^j \mathcal{E}^{-2} \sum_{k \geq j} \sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle f'', \phi_P \rangle|^2.$$

Fixing  $k \geq j$ , we apply Minkowski's inequality to obtain

$$\sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle f'', \phi_P \rangle|^2 \leq \left( \sum_{\substack{K: |K| = 2^{1-k}|J| \\ K \cap 2^{\ell+2}J = \emptyset}} \left( \sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle \mathbb{1}_K f'', \phi_P \rangle|^2 \right)^{1/2} \right)^2$$

where above, we sum over dyadic intervals  $K$  and use the fact that  $f''$  is supported on  $\mathbb{R} \setminus 2^{\ell+5}J$ .

Now note that  $\phi_{P'} = c \exp(2\pi i(c(\omega'_u) - c(\omega_u))) \phi_P$  when  $I = I'$ . Thus if  $\mathcal{P}(I_o)$  is a collection of disjoint multitiles with common time interval  $I_o$ , if  $g$  is supported on  $CI_o$  and if  $P_{I_o}$  is any fixed multitile in  $\mathcal{P}(I_o)$  then Bessel's inequality gives

$$\sum_{P \in \mathcal{P}(I_o)} |\langle g, \phi_P \rangle|^2 \lesssim |I_o| \int |g \phi_{P_{I_o}}|^2 dx.$$

We apply this observation to the inner sums in the previous display and obtain the inequality

$$(33) \quad \sum_{\substack{P: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} |\langle f'', \phi_P \rangle|^2 \lesssim 2^{-k}|J| \left( \sum_{\substack{K: |K| = 2^{1-k}|J| \\ K \cap 2^{\ell+2}J = \emptyset}} \left( \sum_{\substack{I: |I| = 2^{-k}|J| \\ I \subset 2^{\ell+1}J}} \|\mathbb{1}_K f'' \phi_{P_I}\|_{L^2}^2 \right)^{1/2} \right)^2$$

where for each  $I$ ,  $P_I$  is any multitile with time interval  $I$ . Since  $|f''| \leq 1$  the bound (14) yields

$$\|\mathbb{1}_K f'' \phi_{P_I}\|_{L^2}^2 \lesssim \left(1 + \frac{\text{dist}(K, I)}{|I|}\right)^{-N}.$$

Applying it with large  $N$  we see that the right hand side of (33) is  $\lesssim 2^{-(k+\ell)(N-4)} 2^{-k}|J|$ . Summing over  $k \geq j$  we obtain inequality (32). This concludes the proof of (18).

*Proof of (20).* For each  $T \in \mathbf{T}$ , let  $S \equiv S(T)$  be the corresponding  $l$ -overlapping tree from the selection algorithm above and recall

$$\sum_{T \in \mathbf{T}} |I_T| (\mathcal{E}/2)^2 \leq \sum_{P \in \bigcup_{T \in \mathbf{T}} S} |\langle f, \phi_P \rangle|^2.$$

Since  $\mathbf{P} = \bigcup_{T' \in \mathbf{T}'} T'$ , the right side above is dominated by

$$(34) \quad \sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \in C_2 \omega_l}} |\langle f, \phi_P \rangle|^2 + \sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \geq \inf C_3 \omega_u}} |\langle f, \phi_P \rangle|^2 \\ + \sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \sup C_2 \omega_l \leq \xi_{T'} < \inf C_3 \omega_u}} |\langle f, \phi_P \rangle|^2.$$

For each  $T' \in \mathbf{T}'$  the set of tiles  $P \in T' \cap \bigcup_{T \in \mathbf{T}} S(T)$  with the property that  $\xi_{T'} \in C_2\omega_l$  is by definition an  $l$ -overlapping tree. Thus, since  $\mathbf{P}$  has energy bounded by  $\mathcal{E}$ , we estimate the first term in (34)

$$\sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \in C_2\omega_l}} |\langle f, w_u \rangle|^2 \leq \sum_{T' \in \mathbf{T}'} \mathcal{E}^2 |I_{T'}|.$$

For the estimation of the second term in (34) we observe that any fixed multitile forms an  $l$ -overlapping tree (with respect to some top data), and we use the energy bound for fixed tiles. We observe that the rectangles  $\{I \times [\inf C_3\omega_u, \sup C_2\omega_u) : P \in \bigcup_{T \in \mathbf{T}} S\}$  are pairwise disjoint. Indeed if  $P \in S$ ,  $P' \in S$  this follows since  $S$  is  $l$ -overlapping (cf. Observation 4.1). If  $P \in S_{\kappa_1}$  and  $P' \in S_{\kappa_2}$   $\kappa_1 < \kappa_2$  then  $\xi_{S_{\kappa_1}} \geq \xi_{S_{\kappa_2}}$  and an overlap of  $[\inf C_3\omega_u, \sup C_2\omega_u)$  and  $[\inf C_3\omega'_u, \sup C_2\omega'_u)$  would imply that  $P'$  belongs to the maximal tree with the same top data as  $S_{\kappa_1}$ , i.e. this would imply that  $P' \in T_{\kappa_1}$  which is disjoint from  $S_{\kappa_2}$ .

This allows us to estimate for any fixed  $T' \in \mathbf{T}'$

$$\sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \geq \inf C_3\omega_u}} |\langle f, \phi_P \rangle|^2 \leq \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \xi_{T'} \geq \inf C_3\omega_u}} \mathcal{E}^2 |I| \leq \mathcal{E}^2 |I_{T'}|.$$

Now sum over  $T' \in \mathbf{T}'$  and it follows that the middle term in (34) is  $\lesssim \sum_{T' \in \mathbf{T}'} \mathcal{E}^2 |I_{T'}|$  and summing

For the estimation of the third term in (34) we begin with a preliminary observation, also related to the selection of the  $S$ . Suppose  $P \in T' \cap S$ ,  $\tilde{P} \in T' \cap \tilde{S}$  where  $T, \tilde{T} \in \mathbf{T}$  and

$$\xi_{T'} \in [\sup C_2\omega_l, \inf C_3\omega_u) \cap [\sup C_2\tilde{\omega}_l, \inf C_3\tilde{\omega}_u),$$

and suppose  $I \subset \tilde{I}$  and  $P \neq \tilde{P}$ . From (16) we have  $I \subsetneq \tilde{I}$ . We also have  $\inf C_2\tilde{\omega}_l < \sup C_2\omega_l$  since otherwise it would follow that  $\tilde{S}$  was selected prior to  $S$  and hence  $P \in \tilde{T}$  which is impossible. From (15), we have  $\inf C_2\tilde{\omega}_l \geq \sup C_3\omega_l$  and so  $P$  is in the maximal  $l$ -overlapping tree contained in  $T'$  with top data  $(\tilde{I}, \inf C_2\tilde{\omega}_l)$ .

For each  $T' \in \mathbf{T}'$  let  $T''$  be the collection of multitiles  $P \in T' \cap \bigcup_{T \in \mathbf{T}} S$  with  $\xi_{T'} \in [\sup C_2\omega_l, \inf C_3\omega_u)$  and  $I$  maximal among such multitiles. Then

$$\sum_{T' \in \mathbf{T}'} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \sup C_2\omega_l \leq \xi_{T'} < \inf C_3\omega_u}} |\langle f, \phi_P \rangle|^2 \leq \sum_{T' \in \mathbf{T}'} \sum_{P'' \in T''} \sum_{\substack{P \in T' \cap \bigcup_{T \in \mathbf{T}} S \\ \sup C_2\omega_l \leq \xi_{T'} < \inf C_3\omega_u \\ I \subset I''}} |\langle f, \phi_P \rangle|^2$$

Considering the discussion in the preceding paragraph, we may apply the energy bound to the maximal  $l$ -overlapping tree contained in  $T'$  with top data  $(\tilde{I}, \inf C_2\tilde{\omega}_l)$ . Since the  $I''$  are disjoint subintervals of  $I_{T'}$  we see that the right side in the last display is bounded by

$$\sum_{T' \in \mathbf{T}'} \sum_{P'' \in T''} 2\mathcal{E}^2 |I''| \leq \sum_{T' \in \mathbf{T}'} 2\mathcal{E}^2 |I_{T'}|.$$

This completes the proof of (20).  $\square$

The proposition below is for use in tandem with Proposition 4.3.

**Proposition 4.4.** *Let  $\mathbf{P}$  be a collection of multitiles and  $\mu > 0$ . Then, there is a collection of trees  $\mathbf{T}$  such that*

$$(35) \quad \sum_{T \in \mathbf{T}} |I_T| \lesssim \mu^{-r'} |E|$$

and such that

$$\text{density}(\mathbf{P} \setminus \cup_{T \in \mathbf{T}} T) \leq \mu/2.$$

*Proof.* We select trees through an iterative procedure. Suppose that trees  $T_j, T_j^+, T_j^-$  have been chosen for  $j = 1, \dots, k$ . Let

$$\mathbf{P}_k = \mathbf{P} \setminus \bigcup_{j=1}^k T_j \cup T_j^+ \cup T_j^-.$$

If  $\text{density}(\mathbf{P}_k) \leq \mu/2$  then we terminate the procedure and set

$$\mathbf{T} = \{T_1, T_1^+, T_1^-, \dots, T_k, T_k^+, T_k^-\}.$$

Otherwise, we may find a nonempty tree  $T \subset \mathbf{P}_k$  such that

$$(36) \quad \frac{1}{|I_T|} \int_E \left(1 + \frac{|x - c(I_T)|}{|I_T|}\right)^{-4} \sum_{k: \xi_{k-1}(x) \in \omega_T} |a_k(x)|^{r'} dx > (\mu/2)^{r'}.$$

Choose  $T_{k+1} \subset \mathbf{P}_k$  so that  $|I_{T_{k+1}}|$  is maximal among all nonempty trees contained in  $\mathbf{P}_k$  which satisfy (36), and so that  $T_{k+1}$  is the maximal, with respect to inclusion, tree contained in  $\mathbf{P}_k$  with top data  $(I_{T_{k+1}}, \xi_{T_{k+1}})$ . Let  $T_{k+1}^+ \subset \mathbf{P}_k$  be the maximal tree contained in  $\mathbf{P}_k$  with top data  $(I_{T_{k+1}}, \xi_{T_{k+1}} + (C_2 - 1)/(2|I_{T_{k+1}}|))$  and  $T_{k+1}^- \subset \mathbf{P}_k$  be the maximal tree contained in  $\mathbf{P}_k$  with top data  $(I_{T_{k+1}}, \xi_{T_{k+1}} - (C_2 - 1)/(2|I_{T_{k+1}}|))$ . Since each  $T_j$  is nonempty and  $\mathbf{P}$  is finite, this process will eventually stop.

To prove (35), it will suffice to verify

$$(37) \quad \sum_j |I_{T_j}| \lesssim \mu^{-r'} |E|.$$

To this end, we first observe that the tiles  $I_{T_j} \times \omega_{T_j}$  are pairwise disjoint. Indeed, suppose that  $(I_{T_j} \times \omega_{T_j}) \cap (I_{T_{j'}} \times \omega_{T_{j'}}) \neq \emptyset$  and  $j < j'$ . Then, by the first maximality condition, we have  $|I_{T_j}| \geq |I_{T_{j'}}|$  and so  $I_{T_{j'}} \subset I_{T_j}$  and  $|\omega_{T_j}| \leq |\omega_{T_{j'}}|$ . From the latter inequality, it follows that for every  $P \in T_{j'}$ , either  $\omega_{T_j} \subset \omega_m$ ,  $\omega_{T_j^+} \subset \omega_m$ , or  $\omega_{T_j^-} \subset \omega_m$ . Thus,  $T_{j'} \subset T_j \cup T_j^+ \cup T_j^-$  which contradicts the selection algorithm.

Breaking the integral up into pieces and applying a pigeonhole argument, it follows from (36) that for each  $j$  there is a positive integer  $\ell_j$  such that

$$(38) \quad |I_{T_j}| \leq C 2^{-3\ell_j} \mu^{-r'} \int_{E \cap 2^{\ell_j} I_{T_j}} \sum_{k: \xi_{k-1}(x) \in \omega_{T_j}} |a_k(x)|^{r'} dx.$$

For each  $\ell$  we let  $\mathbf{T}^{(\ell)} = \{T_j : \ell_j = \ell\}$  and choose elements of  $\mathbf{T}^{(\ell)}$ :  $T_1^{(\ell)}, T_2^{(\ell)}, \dots$  and subsets of  $\mathbf{T}^{(\ell)}$ :  $\mathbf{T}_1^{(\ell)}, \mathbf{T}_2^{(\ell)}, \dots$  as follows. Suppose  $T_j^{(\ell)}$  and  $\mathbf{T}_j^{(\ell)}$  have been chosen

for  $j = 1, \dots, k$ . If  $\mathbf{T}^{(\ell)} \setminus \bigcup_{j=1}^k \mathbf{T}_j^{(\ell)}$  is empty, then terminate the selection procedure. Otherwise, let  $T_{k+1}^{(\ell)}$  be an element of  $\mathbf{T}^{(\ell)} \setminus \bigcup_{j=1}^k \mathbf{T}_j^{(\ell)}$  with  $|I_{T_{k+1}^{(\ell)}}|$  maximal, and let

$$\mathbf{T}_{k+1}^{(\ell)} = \{T \in \mathbf{T}^{(\ell)} \setminus \bigcup_{j=1}^k \mathbf{T}_j^{(\ell)} : (2^\ell I_T \times \omega_T) \cap (2^\ell I_{T_{k+1}^{(\ell)}} \times \omega_{T_{k+1}^{(\ell)}}) \neq \emptyset\}.$$

By construction,  $\mathbf{T}^{(\ell)} = \bigcup_j \mathbf{T}_j^{(\ell)}$  and so

$$(39) \quad \sum_{T \in \mathbf{T}^{(\ell)}} |I_T| \leq \sum_j \sum_{T \in \mathbf{T}_j^{(\ell)}} |I_T|.$$

Using the fact that the tiles  $I_{T_j} \times \omega_{T_j}$  are pairwise disjoint, and (twice) the fact that  $|I_T| \leq |I_{T_j^{(\ell)}}|$  for every  $T \in \mathbf{T}_j^{(\ell)}$ , we see that for each  $j$

$$\sum_{T \in \mathbf{T}_j^{(\ell)}} |I_T| \lesssim 2^\ell |I_{T_j^{(\ell)}}|.$$

From (38), we thus see that the right side of (39) is dominated by a constant times

$$\begin{aligned} & 2^{-2\ell} \mu^{-r'} \int_E \sum_j \mathbb{1}_{2^\ell I_{T_j^{(\ell)}}}(x) \sum_{k: \xi_{k-1}(x) \in \omega_{T_j^{(\ell)}}} |a_k(x)|^{r'} dx \\ &= 2^{-2\ell} \mu^{-r'} \int_E \sum_k |a_k(x)|^{r'} \# \{j : (x, \xi_{k-1}(x)) \in 2^\ell I_{T_j^{(\ell)}} \times \omega_{T_j^{(\ell)}}\} dx \\ &\leq 2^{-2\ell} \mu^{-r'} |E|. \end{aligned}$$

where we used the disjointness of the rectangles  $2^\ell I_{T_j^{(\ell)}} \times \omega_{T_j^{(\ell)}}$ , and  $\sum_{k=1}^K |a_k(x)|^{r'} \leq 1$ . Summing over  $\ell$ , we obtain (37).  $\square$

## 5. THE TREE ESTIMATE

In this section we prove the basic estimate for the model operators in the special case where the collection of multitiles is a tree. In what follows we use the notation  $\mathcal{V}^r A f(x)$  for the  $r$ -variation of  $k \mapsto A_k f(x)$ , for a given family of operators  $A_k$  indexed by  $k \in \mathbb{N}$ .

An essential tool introduced to harmonic analysis by Bourgain [1] is *Lépingle's inequality* for martingales ([22]). Consider the martingale of dyadic averages

$$\mathbb{E}_k[f](x) = \frac{1}{|I_k(x)|} \int_{I_k(x)} f(y) dy$$

where  $I_k(x)$  is the dyadic interval of length  $2^k$  containing  $x$ . It is a special case of Lépingle's inequality that

$$(40) \quad \|\mathcal{V}^r \mathbb{E}_{(\cdot)}[f]\|_{L^p} \leq C_{p,r} \|f\|_{L^p}$$

whenever  $1 < p < \infty$  and  $r > 2$ . Simple proofs (based on jump inequalities) have been obtained in [1] and [27] (see also [9], [15] for other expositions). Inequality (40) has been extended to various families of convolution operators ([1], [14], [2], [15]). Let



$\psi$  be a Schwartz function on  $\mathbb{R}$  with  $\int \psi = 1$ , for each  $k$  let  $\psi_k = 2^{-k}\psi(2^{-k}\cdot)$  and let  $A_k f(x) = \psi_k * f$ . Then

$$(41) \quad \|\mathcal{V}^r A f\|_{L^p} \leq C'_{p,r} \|f\|_{L^p}, \quad 1 < p < \infty, \quad r > 2,$$

follows from (40); the essential tool is the square function estimate

$$(42) \quad \left\| \left( \sum_{k=-\infty}^{\infty} |A_k f - \mathbb{E}_k[f]|^2 \right)^{1/2} \right\|_p, \quad 1 < p < \infty.$$

For a proof we refer to [14] or [15].

We shall use (41) to prove the following estimate in terms of energy and density. The bound will be applied in Section 7 with  $q = r'$  and  $q = 1$ .

**Proposition 5.1.** *Let  $\mathcal{E} > 0$ ,  $0 < \mu < 3$  and let  $T$  be a tree with energy bounded above by  $\mathcal{E}$  and density bounded above by  $\mu$ . Then, for each  $1 \leq q \leq 2$*

$$(43) \quad \left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}.$$

Furthermore, for  $\ell \geq 0$  we have

$$(44) \quad \left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(\mathbb{R} \setminus 2^\ell I_T)} \lesssim 2^{-\ell(N-10)} \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}.$$

We remark that the bounds above also hold for  $2 < q < \infty$ , but the result for this range of exponents is not needed for our purposes; it requires an additional  $L^p$  estimate for  $\sum_{P \in T} \langle f, \phi_P \rangle \phi_P$ .

*Proof.* We begin with some preliminary reduction. Every tree can be split into an  $l$ -lacunary tree and an  $l$ -overlapping tree and it suffices to prove the asserted estimate for these cases.

If the tree is  $l$ -overlapping we introduce further decompositions. By breaking up  $T$  into a bounded number of subtrees we may and shall assume without loss of generality that for each  $P \in T$ ,

$$(45) \quad \xi_T \in \omega_l + j|\omega_l| \text{ for some integer } j \text{ with } |j| \leq C_2.$$

Moreover we shall assume, for every  $l$ -overlapping tree  $T$ , that either  $\xi_T \leq \inf \omega_l$  for every  $P \in T$  (in which case we refer to  $T$  as  $l^-$ -overlapping) or  $\xi_T > \inf \omega_l$  for every  $P \in T$  (in which case we refer to  $T$  as  $l^+$ -overlapping). Every  $l$ -overlapping tree can be split into an  $l^-$ -overlapping and an  $l^+$ -overlapping tree. For the remainder of the proof, we assume without loss of generality that  $T$  is either  $l$ -lacunary or  $l^+$ -overlapping or  $l^-$ -overlapping, and that for the last two categories property (45) is satisfied.

Let  $\mathbf{J}$  be the collection of dyadic intervals  $J$  which are maximal with respect to the property that  $I \not\subset 3J$  for every  $P \in T$ .

Our first goal is to prove that for each  $J \in \mathbf{J}$

$$(46) \quad \left\| \sum_{P \in T: |I| < C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |J|^{1/q} \left( 1 + \frac{\text{dist}(I_T, J)}{|I_T|} \right)^{-(N-6)}$$

where  $C'' \geq 1$  is a constant to be determined later; we shall see that  $C'' = 8C_1(C_2 - 1)^{-1}$  is an admissible choice.

By Hölder's inequality, we may assume that  $q \geq r'$ . Fix  $P \in T$  with  $|I| \leq C''|J|$ . From the energy bound, we have

$$(47) \quad \|\langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E\|_{L^q(J)} \lesssim \mathcal{E} \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^{-N} \|a_P \mathbb{1}_E\|_{L^q(J)}.$$

From the density bound applied to  $\approx 1/(C_2 - 1)$  nonempty trees, each with top time interval  $I$ , we obtain

$$\frac{1}{|I|} \int_E \left(1 + \frac{|x - c(I)|}{|I|}\right)^{-4} \sum_{k: \xi_{k-1}(x) \in \omega_l} |a_k(x)|^{r'} dx \lesssim \mu^{r'}.$$

Since  $I \not\subset 3J$ , it follows that  $1 + |x - y|/|I| \leq C(1 + \text{dist}(I, J)/|I|)$  for every  $x \in J$  and  $y \in I$ . Thus

$$\|a_P \mathbb{1}_E\|_{L^q(J)}^q \leq \|a_P \mathbb{1}_E\|_{L^{r'}(J)}^{r'} \lesssim \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^4 |I| \mu^{r'}$$

where, above, we use the fact that  $|a_P| \leq 1$ . Thus we can replace (47) by

$$\|\langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E\|_{L^q(J)} \lesssim \mathcal{E} \mu^{r'/q} |I|^{1/q} \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^{-(N-4)}.$$

Summing this estimate and using the fact that  $T$  is a tree, we have

$$\left\| \sum_{P \in T: |I|=2^{-k}|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)} \lesssim \mathcal{E} \mu^{r'/q} (2^{-k}|J|)^{1/q} \left(1 + \frac{\text{dist}(I_T, J)}{|I_T|}\right)^{-(N-6)}$$

and summing over  $k$  gives (46).

We now use (46) to prove (44) for  $\ell \geq 4$ . Indeed, using the maximality of each  $J$ , we see that if  $\ell \geq 4$  and  $J \cap (\mathbb{R} \setminus 2^\ell I_T) \neq \emptyset$  then  $\text{dist}(I_T, J) \geq |J|/2$  and  $|J| \geq 2^{\ell-3}|I_T|$ . It thus follows from (46) that

$$\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)} \lesssim (|I_T|/|J|) (\text{dist}(I_T, J)/|J|)^{-2} \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q} 2^{-\ell(N-10)}$$

whenever  $J \cap (\mathbb{R} \setminus 2^\ell I_T) \neq \emptyset$ . Summing over all  $J$ , we thus obtain (44) for  $\ell \geq 4$ .

It remains to prove

$$\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(16I_T)} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}$$

which, by (46), follows from

$$(48) \quad \left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(16I_T)} \lesssim \mathcal{E} \mu^{\min(1, r'/q)} |I_T|^{1/q}.$$

Again, by Hölder's inequality, we may assume that  $q \geq r'$ . Let

$$\Omega_J = \bigcup_{P \in T: |I| \geq C''|J|} \omega_l.$$

The first step in the proof of (48) will be to demonstrate

$$(49) \quad \int_{J \cap E} \sum_{k: \xi_{k-1}(x) \in \Omega_J} |a_k(x)|^{r'} dx \lesssim \mu^{r'} |J|.$$

By the maximality of  $J$  there is a multitile  $P^* = (I^*, \omega_u^*) \in T$  with  $I^* \subset 3\tilde{J}$  where  $\tilde{J}$  is the dyadic double of  $J$ . This implies that there is a dyadic interval  $J'$  with  $|J| \leq |J'| \leq 4|J|$

and  $\text{dist}(J, J') \leq |J|$  and  $I^* \subset J'$ . We wish to apply the density bound assumption to a tree consisting of the one multitile  $P^*$ , with top interval  $J'$  and suitable choice of the top frequency. We distinguish the cases that  $T$  is  $l^+$ -overlapping,  $l^-$ -overlapping, or  $l$ -lacunary.

If  $T$  is  $l^+$ -overlapping then  $T' = (\{P^*\}, \xi_T, J')$  is a tree. For every  $P \in T$  we have  $\omega_l \subset [\xi_T - \frac{C_1}{2|I|}, \xi_T + \frac{C_1}{2|I|}]$  and thus, with  $|I| \geq C''|J|$ ,  $\omega_l \subset [\xi_T - \frac{C_1}{2C''|J|}, \xi_T + \frac{C_1}{2C''|J|}]$ . Thus with the choice of  $C'' \geq \frac{8C_1}{C_2-1}$  we have  $\Omega_J \subset \omega_{T'}$ .

If  $T$  is  $l^-$ -overlapping then  $T' = (\{P^*\}, \xi_T + \frac{C_2-1}{4|J|}, J')$  is a tree. Using that  $T$  is  $l^-$ -overlapping, we see that  $\omega_l \subset [\xi_T, \xi_T + \frac{C_2}{2C''|J|}]$  for every  $P \in T$  with  $|I| \geq C''|J|$ . Thus, by choosing  $C'' \geq \frac{8C_2}{C_2-1}$ , we have  $\Omega_J \subset \omega_{T'}$ .

If  $T$  is  $l$ -lacunary then  $T' = (\{P^*\}, \xi_T - \frac{C_2-1}{4|J|}, J')$  is a tree. Using that  $T$  is  $l$ -lacunary, we see that  $\omega_l \subset [\xi_T - \frac{C_1}{2C''|J|}, \xi_T]$  for every  $P \in T$  with  $|I| \geq C''|J|$ . Thus, by choosing  $C'' \geq 8\frac{C_1}{C_2-1}$  as in the first case we have  $\Omega_J \subset \omega_{T'}$ .

In any of the three cases, the density bound gives

$$\frac{1}{|J'|} \int_E \left(1 + \frac{|x - c(J')|}{|J'|}\right)^{-4} \sum_{k: \xi_{k-1}(x) \in \omega_{T'}} |a_k(x)|^{r'} dx \leq \mu^{r'}$$

and hence (49).

We now show that if  $T$  is  $l$ -lacunary then (48) follows from (49). We start by observing that for each  $x$  there is at most one integer  $m$  and at most one integer  $k$  such that there exists a  $P \in T$  with  $|I| = 2^m$ ,  $\xi_{k-1}(x) \in \omega_l$ , and  $\xi_k(x) \in \omega_h$ . Indeed, suppose such a  $P$  exists, and  $P' \in T$  with  $|I'| \geq |I|$ . If  $|I| = |I'|$  the uniqueness of  $k$  is obvious. Suppose  $|I'| > |I|$ . Since  $T$  is  $l$ -lacunary, we have  $\inf(\omega'_l) > \sup(\omega_l)$  by (15), and so  $\xi_{k-1}(x) < \inf(\omega'_l)$ . We also have  $\xi_k(x) \geq \inf(\omega_h) > \sup(\omega_T) > \sup(\omega'_l)$  since  $C_2\omega_u \cap \omega_h = \emptyset$  and  $\xi_T \notin C_2\omega'_l$ . It follows that there is no  $k'$  with  $\xi_{k'-1}(x) \in \omega'_l$ .

Now let  $a(x) = a_k(x)$  if there exists an  $m(x)$  as in the previous paragraph with  $2^{m(x)} \geq C''|J|$ , and  $a(x) = 0$  otherwise. We then have

$$\left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)}^q \leq \int_{J \cap E} \left( |a(x)| \sum_{P \in T: |I|=2^{m(x)}} |\langle f, \phi_P \rangle \phi_P(x)| \right)^q dx.$$

From the energy bound (applied to multitiles) and the bound (14) for  $|\phi_P|$ , the right side is bounded by

$$\int_{J \cap E} \left[ |a(x)| \sum_{\substack{P \in T: \\ |I|=2^{m(x)}}} \mathcal{E} |I| w_I(x) \right]^q dx \lesssim \int_{J \cap E} \left[ |a(x)| \sum_{\substack{P \in T: \\ |I|=2^{m(x)}}} \mathcal{E} \left(1 + \frac{|x - c(I)|}{|I|}\right)^{-N} \right]^q dx.$$

Noting that  $\sum_{P \in T: |I|=2^{m(x)}} (1 + |x - c(I)|/|I|)^{-N} \leq C$ , we see that the last display is

$$\lesssim \mathcal{E}^q \int_{J \cap E} |a(x)|^q dx \lesssim \mathcal{E}^q \int_{J \cap E} \left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} |a_k(x)|^{r'} \right)^{q/r'} dx,$$

by our choice of  $a(x)$ . Using (49),  $q \geq r'$ , and the fact that  $\sum |a_k(x)|^{r'} \leq 1$ , the right hand side is  $\lesssim \mathcal{E}^q \mu^{r'} |J|$ . We may now sum over those  $J \in \mathbf{J}$  which satisfy  $J \cap 16I_T \neq \emptyset$  and (48) follows for  $l$ -lacunary trees.

It remains to prove (48) for the case when  $T$  is  $l$ -overlapping and satisfies condition (45). For each  $J \in \mathbf{J}$ , and each  $x \in J \cap E$  we have by Hölder's inequality

$$(50) \quad \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P(x) a_P(x) = \sum_k a_k(x) \sum_{\substack{P \in T: |I| \geq C''|J| \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) \\ \leq \left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} |a_k(x)|^{r'} \right)^{1/r'} \left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} \left| \sum_{\substack{P \in T: |I| \geq C''|J| \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) \right|^r \right)^{1/r}.$$

Now let  $\psi$  be a Schwartz function with  $\widehat{\psi}(\xi) = 1$  for  $|\xi| \leq C_1 + C_3$  and  $\widehat{\psi}(\xi) = 0$  for  $|\xi| \geq 2C_1$ . Define  $\psi_\ell = 2^{-\ell} \psi(2^{-\ell} \cdot)$ , and set

$$e_T(x) = e^{2\pi i \xi_T x}.$$

We will show that for any  $x$  and any  $k$  with  $\xi_{k-1}(x) < \xi_k(x)$ , there exist integers  $\ell_1, \ell_2$  depending on  $x$  and  $k$ , such that  $2^{\ell_1} \geq |J|$  and

$$(51) \quad \sum_{\substack{P \in T: |I| \geq C''|J|, \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) = (e_T(\psi_{\ell_1} - \psi_{\ell_2})) * \left[ \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right](x).$$

From (15) we have, for each  $\ell$  such that  $2^\ell = |I|$  for some multitile  $P$ ,

$$(e_T \psi_\ell) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(x) = \sum_{P \in T: |I| \geq 2^\ell} \langle f, \phi_P \rangle \phi_P(x).$$

Thus, to prove (51) it will suffice to show that there exist integers  $\ell_1$  and  $\ell_2$  such that

$$(52) \quad \{P \in T : |I| \geq C''|J|, \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h\} = \{P \in T : 2^{\ell_1} \leq |I| \leq 2^{\ell_2}\}.$$

Again using (15), we see that for  $P, P' \in T$  with  $|I| < |I'|$  we have  $\inf \omega'_h < \inf \omega_h$ , and if we are in the setting of  $\rho$ -multitiles where  $\rho$  is a 1-index, we have the stronger inequality  $\sup \omega'_h < \inf \omega_h$ . Thus, (52) will follow after finding  $\ell_1$  and  $\ell_2$  with

$$(53) \quad \{P \in T : \xi_{k-1}(x) \in \omega_l\} = \{P \in T : 2^{\ell_1} \leq |I| \leq 2^{\ell_2}\}.$$

Here we use assumption (45). The displayed equation follows when  $|j| > 1$  from the fact that  $\omega_l \cap \omega'_l = \emptyset$  if  $P, P' \in T$  and  $|I| < |I'|$ ; it follows when  $j = 0$  from the fact that the intervals  $\{\omega_l : P \in T\}$  are nested. Finally, when  $j = \pm 1$  it follows from the property that if  $P, P', P'' \in T$ ,  $|I| \leq |I'| \leq |I''|$  and  $\omega_l \cap \omega''_l \neq \emptyset$  then  $\omega'_l \subset \omega'_l \subset \omega_l$ . Thus we have established (53) and consequently (52) and (51).

Using (51), we have

$$\left( \sum_{k: \xi_{k-1}(x) \in \Omega_J} \left| \sum_{\substack{P \in T: |I| \geq C''|J|, \\ \xi_{k-1}(x) \in \omega_l, \xi_k(x) \in \omega_h}} \langle f, \phi_P \rangle \phi_P(x) \right|^r \right)^{1/r} \\ \leq \| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(x) \|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))}$$

where the notation refers to the variation norm with respect to the variable  $k$ , restricted to  $\{k \in \mathbb{Z} : |k| \geq \log_2(|J|)\}$ . For  $\log_2(|J|) \leq k_1 < k_2$ , we have

$$(e_T(\psi_{k_1} - \psi_{k_2})) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P = (e_T \psi_{C+\log_2(|J|)}) * (e_T(\psi_{k_1} - \psi_{k_2})) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P$$

and so, for  $x \in J$

$$\begin{aligned} & \left\| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(x) \right\|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))} \\ & \lesssim \sup_{x \in J} \sup_{R \geq |J|} \frac{2}{|R|} \int_{x-R}^{x+R} \left\| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(y) \right\|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))} dy. \end{aligned}$$

We now integrate the  $q$ th power of the expressions in (50) over  $E \cap J$  and obtain

$$\begin{aligned} & \left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(J)}^q \leq \int_{J \cap E} \left( \sum_{k: \xi_{k-1}(t) \in \Omega_J} |a_k(t)|^{r'} dt \right)^{q/r'} \times \\ & \sup_{x \in J} \sup_{R \geq |J|} \frac{2}{|R|} \int_{x-R}^{x+R} \left\| (e_T \psi_k) * \sum_{P \in T} \langle f, \phi_P \rangle \phi_P(y) \right\|_{V_k^r(\mathbb{Z}^+ + \log_2(|J|))} dy. \\ & \leq \mu^{r'} \int_J \left| \mathcal{M} \left[ \left\| \psi_k * (e_T^{-1} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P) \right\|_{V_k^r} \right](x) \right|^q dx \end{aligned}$$

where  $\mathcal{M}$  is the Hardy-Littlewood maximal operator. For the last inequality we have used  $q \geq r'$ ,  $\sum |a_k(x)|^{r'} \leq 1$  and inequality (49). Summing over  $J \in \mathbf{J}$  gives

$$\begin{aligned} & \left\| \sum_{P \in T: |I| \geq C''|J|} \langle f, \phi_P \rangle \phi_P a_P \mathbb{1}_E \right\|_{L^q(16I_T)}^q \\ & \lesssim \mu^{r'} \left\| \mathcal{M} \left[ \left\| \psi_k * (e_T^{-1} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P) \right\|_{V_k^r} \right](x) \right\|_{L_x^q(16I_T)}^q. \end{aligned}$$

Since  $q \leq 2$ , it follows from Hölder's inequality that the right side above is

$$\lesssim \mu^{r'} |I_T|^{(2-q)/2} \left\| \mathcal{M} \left[ \left\| \psi_k * (e_T^{-1} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P) \right\|_{V_k^r} \right](x) \right\|_{L_x^2(16I_T)}^q.$$

Applying the variation estimate (41) with  $p = 2$  and the  $L^2$  estimate for  $\mathcal{M}$  one sees that the display above is

$$\lesssim \mu^{r'} |I_T|^{(2-q)/2} \left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right\|_{L^2}^q.$$

To finish the proof, it only remains to see that  $\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right\|_{L^2}^2 \lesssim \mathcal{E}^2 |I_T|$ . The left side of this inequality is dominated by

$$\sum_{P \in T} \sum_{P' \in T} |\langle f, \phi_P \rangle| |\langle f, \phi_{P'} \rangle| |\langle \phi_P, \phi_{P'} \rangle| \leq 2 \sum_{P \in T} |\langle f, \phi_P \rangle|^2 \sum_{P' \in T} |\langle \phi_P, \phi_{P'} \rangle|.$$

Since  $T$  is an  $l$ -overlapping tree, we have  $\langle \phi_P, \phi_{P'} \rangle$  unless  $|I| = |I'|$ , in which case, we have  $|\langle \phi_P, \phi_{P'} \rangle| \lesssim (1 + \text{dist}(I, I')/|I|)^{-N}$ . Therefore we obtain the estimate

$$\left\| \sum_{P \in T} \langle f, \phi_P \rangle \phi_P \right\|_{L^2}^2 \lesssim \sum_{P \in T} |\langle f, \phi_P \rangle|^2 \leq C \mathcal{E}^2 |I_T|.$$

This concludes the proof of (48) and thus the proof of the proposition.  $\square$

## 6. TWO AUXILIARY ESTIMATES

Before we give the argument on how to decompose our operators into trees with suitable energy and density bounds we need two auxiliary estimates.

The following proposition can be found in [34], p. 12, or as a special case of a lemma from [12].

**Proposition 6.1.** *Let  $T$  be an  $l$ -overlapping tree. Let  $\lambda > 0$  and  $\Omega_{\lambda,D} = \{\mathcal{M}_D[\mathbb{1}_F] > \lambda\}$  where  $\mathcal{M}_D$  is the maximal dyadic average operator. Then*

$$\frac{1}{|I_T|} \sum_{P \in T: I \not\subset \Omega_{\lambda,D}} |\langle f, \phi_P \rangle|^2 \lesssim \lambda^2.$$

The second auxiliary estimate is the special case of an estimate from [12], but we will provide a proof for convenience.

**Proposition 6.2.** *Let  $\mathbf{P}$  be a finite set of multitiles, and let  $\lambda > 0$ ,  $F \subset \mathbb{R}$ , and  $|f| \leq \mathbb{1}_F$ . Let  $\Omega_\lambda = \{\mathcal{M}[\mathbb{1}_F] > \lambda\}$ . Then*

$$(54) \quad \left\| \sum_{P \in \mathbf{P}: I \subset \Omega} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus \Omega)} \lesssim \frac{|F|}{\lambda^{1/r}}.$$

*Proof.* Fix  $\ell$  and let  $I_\ell \subset \Omega_\lambda$  be a dyadic interval satisfying

$$(55) \quad 2^\ell I_\ell \subset \Omega_\lambda \text{ and } 2^{\ell+1} I_\ell \not\subset \Omega_\lambda.$$

By Minkowski's inequality, we estimate

$$\begin{aligned} & \left( \sum_{P: I = I_\ell} |\langle f, \phi_P \rangle|^2 \right)^{1/2} \\ & \leq \left( \sum_{P: I = I_\ell} |\langle \mathbb{1}_{4I_\ell} f, \phi_P \rangle|^2 \right)^{1/2} + \sum_{j=2}^{\infty} \left( \sum_{P: I = I_\ell} |\langle \mathbb{1}_{2^{j+1}I_\ell \setminus 2^j I_\ell} f, \phi_P \rangle|^2 \right)^{1/2}. \end{aligned}$$

Using orthogonality, the right hand side is bounded by

$$(4|I_\ell|)^{1/2} \|\mathbb{1}_{4I_\ell} f \phi_{P_0}\|_{L^2} + \sum_{j=2}^{\infty} (2^{j+1}|I_\ell|)^{1/2} \|\mathbb{1}_{2^{j+1}I_\ell \setminus 2^j I_\ell} f \phi_{P_0}\|_{L^2}$$

where  $P_0$  is any multitile with  $I = I_\ell$ . Applying the bounds (14) and  $|f| \leq \mathbb{1}_F$ , we see that the last display is

$$\lesssim |F \cap 4I_\ell|^{1/2} + \sum_{j=2}^{\infty} C 2^{-j(N-1)} |F \cap 2^{j+1} I_\ell|^{1/2}.$$

Since  $2^{\ell+1} I_\ell \not\subset \Omega_\lambda$ , we have  $|F \cap 2^{j+1} I_\ell| \leq C 2^{\max(\ell, j)} |I_\ell| \lambda$  for each  $j$ . Thus, the last display is  $\lesssim (2^\ell \lambda |I_\ell|)^{1/2}$  and we have proved

$$\left( \sum_{P: I = I_\ell} |\langle f, \phi_P \rangle|^2 \right)^{1/2} \lesssim (2^\ell \lambda |I_\ell|)^{1/2}.$$

Similarly,

$$\sup_{P: I = I_\ell} |\langle f, \phi_P \rangle| \lesssim 2^\ell \lambda |I_\ell|^{1/2}$$

and so, by interpolation,

$$(56) \quad \left( \sum_{P:I=I_\ell} |\langle f, \phi_P \rangle|^r \right)^{1/r} \lesssim (2^\ell \lambda)^{1/r'} |I_\ell|^{1/2}$$

whenever  $2 \leq r \leq \infty$ . For each  $\xi, I_\ell$  there is at most one  $P \in \mathbf{P}$  with  $\xi \in \omega_I$  and  $I = I_\ell$ . Thus, using the fact that, for each  $x$ ,  $\sum_{k=1}^K |a_k(x)|^{r'} \leq 1$ , we see that

$$\left\| \sum_{P \in \mathbf{P}: I=I_\ell} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus \Omega_\lambda)} \lesssim (2^\ell \lambda)^{1/r'} |I_\ell|^{1/2} \|\phi_{P_0}\|_{L^1(\mathbb{R} \setminus \Omega_\lambda)}$$

where  $P_0$  is any multitile with  $I_0 = I_\ell$ . Using the fact that  $2^\ell I_\ell \subset \Omega_\lambda$ , it follows that the right side above is

$$\lesssim 2^{-\ell(N-2)} \lambda^{1/r'} |I_\ell|.$$

For  $\ell \geq 0$  let  $\mathcal{I}_\ell$  be the set of all dyadic intervals satisfying (55). If  $I \subset \mathcal{I}_\ell$  then for each  $j > 0$  there are at most 2 intervals  $I' \in \mathcal{I}_\ell$  with  $I' \subset I$  and  $|I'| = 2^{-j}|I|$ . By considering the collection of maximal dyadic intervals in  $\mathcal{I}_\ell$ , one sees that

$$\sum_{I \in \mathcal{I}_\ell} |I| \lesssim |\Omega_\lambda|$$

Thus,

$$\left\| \sum_{P \in \mathbf{P}: I \in \mathcal{I}_\ell} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus \Omega_\lambda)} \lesssim 2^{-\ell(N-2)} \lambda^{1/r'} |\Omega_\lambda|.$$

Summing over  $\ell$  and applying the weak-type 1-1 estimate for  $\mathcal{M}$  then gives (54).  $\square$

## 7. CONCLUSION OF THE PROOF

Let  $r > 2$  and  $r' \leq p < \frac{2r}{r-2}$ . We shall conclude the proof of (13), with  $1 \leq |E| \leq 2$ , and thus of Theorem 1.2. It will then suffice, by Chebyshev's inequality, to show

$$(57) \quad \int_{E \setminus G} |S^\rho[f](x)| dx \leq C |F|^{1/p}$$

for any measurable set  $E$  with  $1 \leq |E| \leq 2$ ,  $|f| \leq \mathbb{1}_F$  and some exceptional set  $G = G(E, F)$  with  $|G| \leq 1/4$ .

We shall repeatedly apply Propositions 4.3 and 4.4. By Lemma 4.2 the density of  $\mathbf{P}$  (with respect to the set  $E$  above) and the energy (with respect to  $f$ ) are bounded by a universal constant  $C$ .

We distinguish between the cases  $|F| > 1$  and  $|F| \leq 1$  and first consider the case when  $|F| > 1$ . Repeatedly applying Propositions 4.3 and 4.4 we write  $\mathbf{P}$  as the disjoint union

$$\mathbf{P} = \bigcup_{j \geq 0} \bigcup_{T \in \mathbf{T}_j} T$$

where each  $\mathbf{T}_j$  is a collection of trees  $T$  each of which have energy bounded by  $C 2^{-j/2} |F|^{1/2}$ , density bounded by  $C 2^{-j/r'}$ , and satisfy

$$\sum_{T \in \mathbf{T}_j} |I_T| \lesssim 2^j.$$

For each  $j$  we apply Proposition 4.3 again, this time using (18) and (20) to write

$$\bigcup_{T \in \mathbf{T}_j} T = \bigcup_{k \geq 0} \bigcup_{T \in \mathbf{T}_{j,k}} T$$

where each tree  $T \in \mathbf{T}_{j,k}$  has energy bounded by  $C2^{-(j+k)/2}|F|^{1/2}$ , density bounded by  $C2^{-j/r'}$ , and satisfies

$$(58) \quad \sum_{T \in \mathbf{T}_{j,k}} |I_T| \lesssim 2^j.$$

Moreover, for every  $\ell \geq 0$

$$(59) \quad \left\| \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \right\|_{BMO} \lesssim 2^{2\ell} 2^{j+k} |F|^{-1}.$$

Inequality (58) implies  $\left\| \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \right\|_{L^1} \lesssim 2^{\ell+j}$ , and we may interpolate the  $L^1$  and the  $BMO$  bound. Here we use a standard technique involving the sharp maximal function from §5 in [11]. It follows that for  $1 \leq q < \infty$

$$\left\| \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \right\|_{L^q} \lesssim 2^{j+k+2\ell} |F|^{-1/q'}.$$

Let  $\epsilon > 0$  be small and  $C' > 0$  be large, depending on  $p, q, r$ . For each  $j, k, \ell$  define

$$G_{j,k,\ell} = \left\{ x : \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T} \geq C' |F|^{-1/q'} 2^{(1+\epsilon)(j+k+2\ell)} \right\}.$$

By Chebyshev's inequality, we have

$$|G_{j,k,\ell}| \leq c' 2^{-\epsilon(j+k+2\ell)},$$

so setting  $G = \bigcup_{j,k,\ell \geq 0} G_{j,k,\ell}$  we have  $|G| \leq 1/4$ .

Applying Minkowski's inequality gives

$$\begin{aligned} & \left\| \mathbb{1}_E \sum_{P \in \mathbf{P}} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G)} \\ & \leq \sum_{j,k \geq 0} \left( \left\| \mathbb{1}_E \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{I_T} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G_{j,k,0})} \right. \\ & \quad \left. + \sum_{\ell \geq 1} \left\| \mathbb{1}_E \sum_{T \in \mathbf{T}_{j,k}} \mathbb{1}_{2^\ell I_T \setminus 2^{\ell-1} I_T} \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G_{j,k,\ell})} \right). \end{aligned}$$

From Hölder's inequality, Fubini's theorem, and the definition of  $G_{j,k,\ell}$ , it follows that the right side above is  $\lesssim (S_1 + S_2)$  where

$$S_1 = \sum_{j,k \geq 0} |F|^{-1/(q'r)} 2^{(1+\epsilon)(j+k)/r} \left( \sum_{T \in \mathbf{T}_{j,k}} \left\| \mathbb{1}_E \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^{r'}(\mathbb{R})}^{r'} \right)^{1/r'}$$

and

$$S_2 = \sum_{\substack{j,k \geq 0 \\ \ell \geq 1}} |F|^{-1/(q'r)} 2^{(1+\epsilon)(j+k+2\ell)/r} \left( \sum_{T \in \mathbf{T}_{j,k}} \left\| \mathbb{1}_E \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^{r'}(\mathbb{R} \setminus 2^{\ell-1} I_T)}^{r'} \right)^{1/r'}$$



Applying Proposition 5.1 with the energy and density bounds for trees  $T \in \mathbf{T}_{j,k}$ , we see that

$$\begin{aligned} S_2 &\lesssim \sum_{\substack{j,k \geq 0 \\ \ell \geq 1}} |F|^{-1/(q'r)} 2^{(1+\epsilon)(j+k+2\ell)/r} 2^{-\ell(N-10)} 2^{-(j+k)/2} |F|^{1/2} 2^{-j/r'} \left( \sum_{T \in \mathbf{T}_{j,k}} |I_T| \right)^{1/r'} \\ &\lesssim \sum_{\substack{j,k \geq 0 \\ \ell \geq 1}} 2^{(j+k)((1+\epsilon)(2/r)-1)/2} 2^{-\ell(N-14)} |F|^{1/2-1/(q'r)}. \end{aligned}$$

Choosing  $\epsilon$  small enough and  $q$  large enough so that

$$(1+\epsilon)\frac{2}{r} - 1 < 0 \text{ and } \frac{1}{2} - \frac{1}{q'r} < \frac{1}{p}$$

we have  $S_2 \lesssim |F|^{1/p}$ . We similarly obtain  $S_1 \lesssim |F|^{1/p}$ , thus giving (57).

We will finish by proving (57) for  $|F| \leq 1$ . Here, we let

$$G = \{x : \mathcal{M}[\mathbb{1}_F](x) > C''|F|\}$$

where  $\mathcal{M}$  is the Hardy-Littlewood maximal operator and  $C''$  is chosen large enough so that the weak-type 1-1 estimate for  $\mathcal{M}$  guarantees  $|G| \leq 1/4$ .

By Proposition 6.2 and the fact that  $p \geq r'$ , it will remain to show that

$$(60) \quad \left\| \mathbb{1}_E \sum_{P \in \mathbf{P}'} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1(\mathbb{R} \setminus G)} \lesssim |F|^{1/p}.$$

where  $\mathbf{P}' = \{P \in \mathbf{P} : I \not\subset G\}$ .

Finally, it follows from Proposition 6.1 that the energy of  $\mathbf{P}'$  is bounded above by  $C|F|$ . Repeatedly applying Propositions 4.3 and 4.4 we write  $\mathbf{P}'$  as the disjoint union

$$\mathbf{P}' = \bigcup_{j \geq 0} \bigcup_{T \in \mathbf{T}_j} T$$

where each  $\mathbf{T}_j$  is a collection of trees  $T$  each of which have energy bounded by  $C_0 2^{-j/2} |F|^{1/2}$ , density bounded by  $C_0 2^{-j/r'}$ , and satisfy

$$\sum_{T \in \mathbf{T}_j} |I_T| \lesssim 2^j.$$

We then have

$$\left\| \mathbb{1}_E \sum_{P \in \mathbf{P}'} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1} \leq \sum_{j \geq 0} \sum_{T \in \mathbf{T}_j} \left\| \mathbb{1}_E \sum_{P \in T} \langle f, \phi_P \rangle \phi_P a_P \right\|_{L^1}.$$

Applying Proposition 5.1, we see that the right side above is

$$\lesssim \sum_{j \geq 0} \sum_{T \in \mathbf{T}_j} \min(2^{-j/2} |F|^{1/2}, |F|) 2^{-j/r'} |I_T| \lesssim \sum_{j \geq 0} 2^{j/r} \min(2^{-j/2} |F|^{1/2}, |F|).$$

Summing over  $j$ , we see that the right side above is  $\lesssim |F|^{1/r'}$ . This finishes the proof, since  $p \geq r'$ .

## A. TRANSFERENCE

In this section we show how to obtain Theorem 1.1 from Theorem 1.2. We employ arguments from chapter VII in the monograph by Stein and Weiss [32] in their proof of De Leeuw's transference result ([21]). The following limiting relation is used:

**Lemma A.1.** ([32], p. 261). *Let  $m \in L^\infty(\mathbb{R}^d)$  with the property that every  $k \in \mathbb{Z}^d$  is a Lebesgue point of  $m$ . Define a convolution operator  $T$  on  $L^2(\mathbb{R}^d)$  by the Fourier transform identity  $\widehat{Tf}(\xi) = m(\xi)\widehat{f}(\xi)$  and a convolution operator on  $L^2(\mathbb{T}^d)$  by the relation  $[\widehat{Tf}]_k = m(k)f_k$  for the Fourier coefficients. Let, for  $x \in \mathbb{R}$ ,*

$$w(x) = e^{-\pi x^2} \text{ and } w_R(x) = R^{-d}w(x/R).$$

*Then, for all trigonometric polynomials  $P$  and  $Q$  (extended as 1-periodic functions in every variable) we have*

$$(61) \quad \int_{[0,1]^d} \widehat{T}[P](x)Q(x) dx = \lim_{R \rightarrow \infty} \int_{\mathbb{R}^d} T[Pw_{R/p'}](x)Q(x)w_{R/p}(x) dx.$$

We also need the following elementary fact on Lorentz spaces.

**Lemma A.2.** *Let  $f \in L^{p,q}(\mathbb{T}^d)$  and extend  $f$  to a function  $f^{\text{per}}$  on  $\mathbb{R}^d$  which is 1-periodic in every variable. Let  $L > d/p$  and let  $w$  be a measurable function satisfying  $|w(x)| \leq (1 + |x|)^{-L}$ . Let  $w_R(x) = R^{-d}w(R^{-1}x)$ . Then*

$$\sup_{R \geq 1} \|f^{\text{per}}w_R\|_{L^{p,q}(\mathbb{R}^d)} \leq C_{p,q} \|f\|_{L^{p,q}(\mathbb{T}^d)}$$

*Proof.* We first assume  $p = q$ . Let  $Q_0 := [-\frac{1}{2}, \frac{1}{2}]^d$ . If  $N \in \mathbb{N}$  and  $N \leq R \leq N+1$  then  $w_R(x) \approx N^{-d}(1 + |n|/N)^{-Lp}$  for  $x \in n + Q_0$  and by the periodicity we can estimate  $\|f^{\text{per}}w_R\|_p^p$  by  $C \sum_{n \in \mathbb{Z}^d} N^{-d}(1 + |n|/N)^{-Lp} \|f\|_{L^p(Q_0)}^p$  which is  $\lesssim \|f\|_{L^p(Q_0)}^p$  since  $Lp > d$ . For fixed  $L$  we apply real interpolation in the range  $p < L/d$  and obtain the Lorentz space result.  $\square$

*Proof that Theorem 1.2 implies Theorem 1.1.* We shall assume that  $1 < p < \infty$ ,  $1 \leq q_1 < \infty$ , and  $q_1 \leq q_2 \leq \infty$  and prove that the  $L^{p,q_1}(\mathbb{R}) \rightarrow L^{p,q_2}(\mathbb{R}; V^r)$  for the partial sum operator  $\mathcal{S}$  on the real line implies the corresponding result on the torus, i.e.

$$(62) \quad \left\| \sup_K \sup_{0 \leq n_1 \leq \dots \leq n_K} \left( \sum_{i=1}^{K-1} |S_{n_{i+1}}f - S_{n_i}f|^r \right)^{1/r} \right\|_{L^{p,q_2}(\mathbb{T})} \lesssim \|f\|_{L^{p,q_1}(\mathbb{T})}$$

By two applications of the monotone convergence theorem it suffices to show for fixed  $K \in \mathbb{N}$  with  $K \geq 2$ , and fixed  $M \in \mathbb{N}$  that

$$(63) \quad \left\| \sup_{0 \leq n_1 \leq \dots \leq n_K \leq M} \left( \sum_{i=1}^{K-1} |S_{n_{i+1}}f - S_{n_i}f|^r \right)^{1/r} \right\|_{L^{p,q_2}(\mathbb{T})} \leq C \|f\|_{L^{p,q_1}(\mathbb{T})}$$

where  $C$  does not depend on  $M$  and  $K$ .

For  $\vec{n} = (n_1, \dots, n_K) \in \mathbb{N}_0^K$ ,  $1 \leq i \leq K-1$  define  $\mathcal{T}f(x, \vec{n}, i) = S_{n_{i+1}}f(x) - S_{n_i}f(x)$  if  $n_1 \leq \dots \leq n_K$  and  $\mathcal{T}f(x, \vec{n}, i) = 0$  otherwise. Then the inequality (63) just says that  $\mathcal{T}$  is bounded from  $L^{p,q_1}$  to  $L^{p,q_2}(\ell^\infty(\ell^r))$  where the  $\ell^\infty$  norm is taken for functions on the

finite set  $\{1, \dots, M\}^K$  and the  $\ell^r$  norm is for functions on  $\{1, \dots, K-1\}$ . By duality (63) follows from the  $L^{p', q'_2}(\ell^1(\ell^{r'})) \rightarrow L^{p', q'_1}$  inequality for the adjoint operator  $\mathcal{T}^*$ , i.e. from the inequality

$$(64) \quad \left| \int_0^1 \sum_{0 \leq n_1 \leq \dots \leq n_K \leq M} \sum_{i=1}^{K-1} [S_{n_{i+1}} f_{\vec{n}, i}(x) - S_{n_i} f_{\vec{n}, i}(x)] Q(x) dx \right| \\ \lesssim \left\| \sum_{\vec{n}} \left( \sum_{i=1}^{K-1} |f_{\vec{n}, i}|^{r'} \right)^{1/r'} \right\|_{L^{p', q'_2}(\mathbb{T})} \|Q\|_{L^{p, q_1}(\mathbb{T})}.$$

We fix an irrational number  $\lambda$  in  $(0, 1)$ , say  $\lambda = 1/\sqrt{2}$ . We then define “partial sum operators” for Fourier integrals by  $\widehat{\mathfrak{S}_t f}(\xi) = \chi_{[-\lambda, \lambda]}(\xi/t) \widehat{f}(\xi)$  and a corresponding partial sum operator  $\widetilde{\mathfrak{S}}_t$  on Fourier series by letting the  $k$ th Fourier coefficient of  $\widetilde{\mathfrak{S}}_t f$  be equal to  $\chi_{[-\lambda, \lambda]}(k/t) \widehat{f}_k$ . We define a function  $\nu : \mathbb{N}_0 \rightarrow \mathbb{N}_0$  as follows: set  $\nu(0) = 0$  and for  $n > 0$  let  $\nu(n)$  be the smallest positive integer  $\nu$  for which  $\lambda\nu > n$ . Notice that then

$$(65) \quad S_{n_{i+1}} f_{\vec{n}, i}(x) - S_{n_i} f_{\vec{n}, i}(x) = \widetilde{\mathfrak{S}}_{\nu(n_{i+1})} f_{\vec{n}, i} - \widetilde{\mathfrak{S}}_{\nu(n_i)} f_{\vec{n}, i}.$$

Now in order to prove (64) it clearly suffices to verify it for the case that the function  $f_{\vec{n}, i}$  and  $Q$  are trigonometric polynomials. The multipliers corresponding to  $\mathfrak{S}_t$  are continuous at every integer. Thus by (61) (applied with  $d = 1$ ) and (65) we see that (64) is implied by

$$(66) \quad \left| \sum_{0 \leq n_1 \leq \dots \leq n_K \leq M} \sum_{i=1}^{K-1} \int_0^1 [\mathfrak{S}_{\nu(n_{i+1})}(f_{\vec{n}, i} w_{R/p'}) - \mathfrak{S}_{\nu(n_i)}(f_{\vec{n}, i} w_{R/p'})] Q w_{R/p} dx \right| \\ \lesssim \left\| \sum_{\vec{n}} \left( \sum_{i=1}^{K-1} |f_{\vec{n}, i}|^{r'} \right)^{1/r'} \right\|_{L^{p', q'_2}(\mathbb{T})} \|Q\|_{L^{p, q_1}(\mathbb{T})}.$$

for sufficiently large  $R$ .

Now notice that  $\mathfrak{S}_t = \mathcal{S}_{\lambda t} - \mathcal{S}_{-\lambda t}$  so that the assumed  $L^{p, q_1}(\mathbb{R}) \rightarrow L^{p, q_2}(V^r, \mathbb{R})$  boundedness for the family  $\{\mathcal{S}_t\}$  implies the analogous statement for the family  $\{\mathfrak{S}_t\}$ . We run the duality argument in the reverse direction (now for functions defined on  $\mathbb{R}$ ) and deduce

$$\left| \int_{\mathbb{R}} \sum_{0 \leq n_1 \leq \dots \leq n_K \leq M} \sum_{i=1}^{K-1} \left( \mathfrak{S}_{\nu(n_{i+1})}[f_{\vec{n}, i} w_{R/p'}] - \mathfrak{S}_{\nu(n_i)}[f_{\vec{n}, i} w_{R/p'}] \right) Q w_{R/p} dx \right| \\ \lesssim \left\| \sum_{\vec{n}} \left( \sum_{i=1}^{K-1} |f_{\vec{n}, i}|^{r'} \right)^{1/r'} w_{R/p'} \right\|_{L^{p', q'_2}(\mathbb{R})} \|Q w_{R/p}\|_{L^{p, q_1}(\mathbb{R})}.$$

By Lemma A.2 the right hand side of this inequality is for  $R \geq \max\{p, p'\}$  bounded by the right hand side of (66). Thus we have established inequality (66) and this concludes the proof.  $\square$

## B. A VARIATIONAL MENSHOV-PALEY-ZYGMUND THEOREM

For  $\xi, x \in \mathbb{R}$  let

$$\mathcal{C}[f](\xi, x) = \int_{-\infty}^x e^{-2\pi i \xi x'} f(x') dx'.$$

Menshov, Paley, and Zygmund extended the Hausdorff-Young inequality by proving a version of the bound

$$(67) \quad \|\mathcal{C}[f]\|_{L_{\xi}^{p'}(L_x^{\infty})} \leq C_p \|f\|_{L^p(\mathbb{R})}$$

for  $1 \leq p < 2$ . The bound at  $p = 2$  is a special case of the much more difficult maximal inequality for the partial sum operator of proved by Carleson and Hunt. Interpolating Theorem 1.2 at  $p = 2$  with a trivial estimate at  $p = 1$ , one obtains the following strengthened version of (67)

$$(68) \quad \|\mathcal{C}[f]\|_{L_{\xi}^{p'}(V_x^r)} \leq C_{p,r} \|f\|_{L^p(\mathbb{R})}$$

for  $1 \leq p \leq 2$  and  $r > p$ . It follows from the same arguments given in Section 2 that this range of  $r$  is the best possible. Our interest in this variational bound primarily stems from the fact, which will be proven in Appendix C, that it may be transferred, when  $r < 2$ , to give a corresponding estimate for certain nonlinear Fourier summation operators. The purpose of the present appendix is to give an easier alternate proof of (68) when  $p < 2$ . Note that Pisier and Xu [27] have proved closely related  $L^p \rightarrow L^2(V^p)$  inequalities for orthonormal systems of (not necessarily bounded) functions on an arbitrary measure space.

A now-famous lemma of Christ and Kiselev [5] asserts that if an integral operator

$$Tf(x) = \int_{\mathbb{R}} K(x, y) f(y) dy$$

is bounded from  $L^p(\mathbb{R})$  to  $L^q(X)$  for some measure space  $X$  and some  $q > p$ , thus

$$\|Tf\|_{L^q(X)} \leq A \|f\|_{L^p(\mathbb{R})},$$

then automatically the maximal function

$$T_*f(x) = \sup_{N \in \mathbb{R}} \left| \int_{y < N} K(x, y) f(y) dy \right|$$

is also bounded from  $L^p(\mathbb{R})$  to  $L^q(X)$ , with a slightly larger constant. Another way to phrase this is as follows. If we define the partial integrals

$$T_{\leq}f(x, N) = \int_{y < N} K(x, y) f(y) dy$$

then we have

$$(69) \quad \|T_{\leq}f\|_{L_x^q(L_N^{\infty})} \leq C_{p,q} A \|f\|_{L^p(\mathbb{R})}.$$

As was observed by Christ and Kiselev, this may be applied in conjunction with the Hausdorff-Young inequality to obtain (67) for  $p < 2$ .

The  $L_N^{\infty}$  norm can also be interpreted as the  $V_N^{\infty}$  norm, and we will now see that  $V^{\infty}$  can be replaced by  $V^r$  for  $r > p$ , thus giving (68) from the Hausdorff-Young inequality.

**Lemma B.1.** *Under the same assumptions, we have*

$$\|T_{\leq} f\|_{L_x^q(V_N^r)} \leq C_{p,q,r} A \|f\|_{L^p(\mathbb{R})}$$

for any  $r > p$ .

*Proof.* This follows by an adaption of the argument by Christ and Kiselev, or by the following argument. Without loss of generality we may take  $r < q$ , in particular  $r < \infty$ . We use a bootstrap argument. Let us make the *a priori* assumption that

$$(70) \quad \|T_{\leq} f\|_{L_x^q(V_N^r)} \leq B A \|f\|_{L^p(\mathbb{R})}$$

for some constant  $0 < B < \infty$ ; this can be accomplished for instance by truncating the kernel  $K$  appropriately. We will show that this *a priori* bound automatically implies the bound

$$(71) \quad \|T_{\leq} f\|_{L_x^q(V_N^r)} \leq (2^{1/r-1/p} B A + C_{p,q,r} A) \|f\|_{L^p(\mathbb{R})}$$

for some  $C_{p,q,r} > 0$ . This implies that the best bound  $B$  in the above inequality will necessarily obey the inequality

$$B \leq 2^{1/r-1/p} B + C_{p,q,r};$$

since  $r > p$ , this implies  $B \leq C'_{p,q,r}$  for some finite  $C'_{p,q,r}$ , and the claim follows.

It remains to deduce (71) from (70). Fix  $f$ ; we may normalize  $\|f\|_{L^p(\mathbb{R})} = 1$ . We find a partition point  $N_0$  in the real line which halves the  $L^p$  norm of  $f$ :

$$\int_{-\infty}^{N_0} |f(y)|^p dy = \int_{N_0}^{+\infty} |f(y)|^p dy = \frac{1}{2}.$$

Write  $f_-(y) = f(y) \mathbb{1}_{(-\infty, N_0]}(y)$  and  $f_+(y) = f(y) \mathbb{1}_{[N_0, +\infty)}(y)$ , thus  $\|f_-\|_{L^p(\mathbb{R})} = 2^{-1/p}$  and  $\|f_+\|_{L^p(\mathbb{R})} = 2^{-1/p}$ . We observe that

$$T_{\leq} f(x, N) = \begin{cases} T_{\leq} f_-(x, N) & \text{when } N \leq N_0 \\ T_{\leq} f_-(x, N) + T_{\leq} f_+(x, N) & \text{when } N > N_0 \end{cases}$$

Furthermore,  $T_{\leq} f_-(x, \cdot)$  and  $T_{\leq} f_+(x, \cdot)$  are bounded in  $L^\infty$  norm by  $O(T_* f(x))$ . Thus we have

$$\|T_{\leq} f(x, \cdot)\|_{V_N^r} \leq (\|T_{\leq} f_-(x, \cdot)\|_{V_N^r}^r + \|T_{\leq} f_+(x, \cdot)\|_{V_N^r}^r)^{1/r} + O(T_* f(x)).$$

(The  $O(T_* f(x))$  error comes because the partition used to define  $\|T_{\leq} f(x, \cdot)\|_{V_N^r}$  may have one interval which straddles  $N_0$ ). We take  $L^q$  norms of both sides to obtain

$$\|T_{\leq} f\|_{L_x^q V_N^r} \leq \|(\|T_{\leq} f_-(x, \cdot)\|_{V_N^r}^r + \|T_{\leq} f_+(x, \cdot)\|_{V_N^r}^r)^{1/r}\|_{L_x^q} + O(\|T_* f\|_{L_x^q}).$$

The error term is at most  $C_{p,q} A$  by the ordinary Christ-Kiselev lemma. For the main term, we take advantage of the fact that  $r < q$  to interchange the  $l^r$  and  $L^q$  norms, thus obtaining

$$\|T_{\leq} f\|_{L_x^q V_N^r} \leq (\|T_{\leq} f_-\|_{L_x^q V_N^r}^r + \|T_{\leq} f_+\|_{L_x^q V_N^r}^r)^{1/r} + O(C_{p,q} A).$$

By inductive hypothesis we thus have

$$\|T_{\leq} f\|_{L_x^q V_N^r} \leq ((2^{-1/p} B A)^r + (2^{-1/p} B A)^r)^{1/r} + O(C_{p,q} A),$$

and the claim follows. □

## C. VARIATION NORMS ON LIE GROUPS

In this appendix, we will show that certain  $r$ -variation norms for curves on Lie groups can be controlled by the corresponding variation norms of their “traces” on the Lie algebra as long as  $r < 2$ . This follows from work of Terry Lyons [23]; we present a self contained proof in this appendix. Combining this fact with the variational Menshov-Paley-Zygmund theorem of Appendix B, we rederive the Christ-Kiselev theorem on the pointwise convergence of the nonlinear Fourier summation operator for  $L^p(\mathbb{R})$  functions,  $1 \leq p < 2$ .

Let  $G$  be a connected finite-dimensional Lie group with Lie algebra  $\mathfrak{g}$ . We give  $\mathfrak{g}$  any norm  $\|\cdot\|_{\mathfrak{g}}$ , and push forward this norm using left multiplication by the Lie group to define a norm  $\|x\|_{T_g G} = \|g^{-1}x\|_{\mathfrak{g}}$  on each tangent space  $T_g G$  of the group. Observe that this norm structure is preserved under left group multiplication.

We can now define the *length*  $|\gamma|$  of a continuously differentiable path  $\gamma : [a, b] \rightarrow G$  by the usual formula

$$|\gamma| = \int_a^b \|\gamma'(t)\|_{T_{\gamma(t)} G} dt.$$

Observe that this notion of length is invariant under left group multiplication, and also under reparameterization of the path  $\gamma$ .

From this notion of length, we can define a metric  $d(g, g')$  on  $G$  as

$$d(g, g') = \inf_{\gamma: \gamma(a)=g, \gamma(b)=g'} |\gamma|$$

where  $\gamma$  ranges over all differentiable paths from  $g$  to  $g'$ . It is easy to see that this does indeed give a metric on  $G$ .

Integral curves of left invariant vectorfields need not be geodesic for this metric [28], but the length of a short segment of such an integral curve is within a quadratically small error of the distance between the two endpoints. This is the content of the following lemma:

**Lemma C.1.** *If  $x \in \mathfrak{g}$  is such that  $\|x\|_{\mathfrak{g}} \leq \epsilon$  for some sufficiently small  $\epsilon$ , then  $d(1, \exp(x)) = \|x\|_{\mathfrak{g}} + O(\|x\|_{\mathfrak{g}}^2)$ .*

*Proof.* By considering the exponential curve  $\gamma : [0, 1] \rightarrow G$  defined by  $\gamma(t) := \exp(tx)$  we obtain the upper bound  $d(1, \exp(x)) \leq \|x\|_{\mathfrak{g}}$ . Now consider any competitor curve  $\tilde{\gamma} : [0, 1] \rightarrow G$  from 1 to  $\exp(x)$  which has shorter length than  $\|x\|_{\mathfrak{g}}$ . We write  $\tilde{\gamma}(t) = \exp(f(t))$  for some smooth curve  $f : [0, 1] \rightarrow \mathfrak{g}$  from 1 to  $x$ ; this is well-defined if  $\epsilon$  is small enough.

There are two cases. First suppose that  $f$  stays inside the ball  $\{y : \|y\|_{\mathfrak{g}} \leq 2\|x\|_{\mathfrak{g}}\}$ . Then from Taylor expansion we see that

$$\|\tilde{\gamma}'(t)\|_{T_{\gamma(t)} G} = (1 + O(\|x\|_{\mathfrak{g}})) \|f'(t)\|_{\mathfrak{g}}$$

and hence

$$|\tilde{\gamma}| = (1 + O(\|x\|_{\mathfrak{g}})) \int_0^1 \|f'(t)\|_{\mathfrak{g}} dt.$$

But from Minkowski's inequality one has  $\int_0^1 \|f'(t)\|_{\mathfrak{g}} dt \geq \|x\|_{\mathfrak{g}}$ , and the claim follows.

Now suppose instead that  $f$  leaves this ball. Let  $0 < t_0 < 1$  be the first time at which this occurs. Then the above argument gives

$$|\tilde{\gamma}| \geq (1 + O(\|x\|_{\mathfrak{g}})) \int_0^{t_0} \|f'(t)\|_{\mathfrak{g}} dt.$$

By Minkowski's inequality one has  $\int_0^{t_0} \|f'(t)\|_{\mathfrak{g}} dt \geq 2\|x\|_{\mathfrak{g}}$ , and this gives a contradiction to  $|\tilde{\gamma}| \leq \|x\|_{\mathfrak{g}} \leq \epsilon$  if  $\epsilon$  is sufficiently small. The claim follows.  $\square$

Given any continuous path  $\gamma : [a, b] \rightarrow G$  and  $1 \leq r < \infty$ , we define the  $r$ -variation  $\|\gamma\|_{V^r}$  of  $\gamma$  to be the quantity

$$\|\gamma\|_{V^r} = \sup_{a=t_0 < t_1 < \dots < t_n=b} \left( \sum_{j=0}^{n-1} d(\gamma(t_{j+1}), \gamma(t_j))^r \right)^{1/r}$$

where the infimum ranges over all partitions of  $[a, b]$  by finitely many times  $a = t_0, t_1, \dots, t_n = b$ . We can extend this to the  $r = \infty$  case in the usual manner as

$$\|\gamma\|_{V^\infty} = \sup_{a=t_0 < t_1 < \dots < t_n=b} \sup_{0 \leq j \leq n-1} d(\gamma(t_{j+1}), \gamma(t_j)),$$

and indeed it is clear that the  $V^\infty$  norm of  $\gamma$  is simply the diameter of the range of  $\gamma$ . The  $V^1$  norm of  $\gamma$  is finite precisely when  $\gamma$  is rectifiable, and when  $\gamma$  is differentiable it corresponds exactly with the length  $|\gamma|$  of  $\gamma$  defined earlier. It is easy to see the monotonicity property

$$\|\gamma\|_{V^p} \leq \|\gamma\|_{V^r} \text{ whenever } 1 \leq r \leq p \leq \infty$$

and the triangle inequalities

$$(\|\gamma_1\|_{V^r}^r + \|\gamma_2\|_{V^r}^r)^{1/r} \leq \|\gamma_1 + \gamma_2\|_{V^r} \leq \|\gamma_1\|_{V^r} + \|\gamma_2\|_{V^r}$$

where  $\gamma_1 + \gamma_2$  is the concatenation of  $\gamma_1$  and  $\gamma_2$ . A key fact about the  $V^r$  norms is that they can be subdivided:

**Lemma C.2.** *Let  $\gamma : [a, b] \rightarrow G$  be a continuously differentiable curve with finite  $V^r$  norm. Then there exists a decomposition  $\gamma = \gamma_1 + \gamma_2$  of the curve into two sub-curves such that*

$$\|\gamma_1\|_{V^r}, \|\gamma_2\|_{V^r} \leq 2^{-1/r} \|\gamma\|_{V^r}.$$

*Proof.* Let  $t_* = \sup\{t \in [a, b] : \|\gamma|_{[a, t]}\|_{V^r} \leq 2^{-1/r} \|\gamma\|_{V^r}\}$ . Letting  $\gamma_1 = \gamma|_{[a, t_*]}$  we have  $\|\gamma_1\|_{V^r} = 2^{-1/r} \|\gamma\|_{V^r}$ . The bound for  $\gamma_2 = \gamma|_{[t_*, b]}$  follows from the left triangle inequality above.  $\square$

Given a continuously differentiable curve  $\gamma : [a, b] \rightarrow G$ , we can define its *left trace*  $\gamma_l : [a, b] \rightarrow \mathfrak{g}$  by the formula

$$\gamma_l(t) = \int_a^t \gamma(s)^{-1} \gamma'(s) ds$$

Note that the trace is also a continuously differentiable curve, but taking values now in the Lie algebra  $\mathfrak{g}$  instead of  $G$ . Clearly  $\gamma_l$  is determined uniquely from  $\gamma$ . The converse is also true after specifying the initial point  $\gamma(a)$  of  $\gamma$ , since  $\gamma$  can then be recovered by solving the ordinary differential equation

$$(72) \quad \gamma'(t) = \gamma(t) \gamma_l'(t).$$

This equation is fundamental in the theory of eigenfunctions of a one-dimensional Schrödinger or Dirac operator, or equivalently in the study of the nonlinear Fourier transform; see, for example, [33], [25] for a full discussion. Basically for a fixed potential  $f(t)$  and a frequency  $k$ , the nonlinear Fourier transform traces out a curve  $\gamma(t)$  (depending on  $k$ ) taking values in a Lie group (e.g.  $SU(1, 1)$ ), and the corresponding left trace is essentially the ordinary linear Fourier transform.

It is easy to see that these curves have the same length (i.e. they have the same  $V^1$  norm):

$$(73) \quad |\gamma| = |\gamma_l|.$$

We now show that something similar is true for the  $V^r$  norms provided that  $r < 2$ .

**Lemma C.3.** *Let  $1 \leq r < 2$ , let  $G$  be a connected finite-dimensional Lie group, and let  $\|\cdot\|_{\mathfrak{g}}$  be a norm on the Lie algebra of  $G$ . Then there exist a constant  $C > 0$  depending only on these above quantities, such that for all smooth curves  $\gamma : [a, b] \rightarrow G$ , we have*

$$(74) \quad \|\gamma\|_{V^r} \leq \|\gamma_l\|_{V^r} + C \min(\|\gamma_l\|_{V^r}^2, \|\gamma_l\|_{V^r}^r)$$

and

$$(75) \quad \|\gamma_l\|_{V^r} \leq \|\gamma\|_{V^r} + C \min(\|\gamma\|_{V^r}^2, \|\gamma\|_{V^r}^r).$$

An analogous result holds for the *right trace*,  $\int_a^t \gamma'(s) \gamma(s)^{-1} ds$ , once the left-invariant norm on  $T_g G$  is replaced by a right-invariant norm.

*Proof.* We may take  $r > 1$  since the claim is already known for  $r = 1$  thanks to (73).

It shall suffice to prove the existence of a small  $\delta > 0$  such that we have the estimate

$$(76) \quad \|\gamma\|_{V^r} = \|\gamma_l\|_{V^r} + O(\|\gamma_l\|_{V^r}^2)$$

whenever  $\|\gamma_l\|_{V^r} \leq \delta$ , and similarly

$$(77) \quad \|\gamma_l\|_{V^r} = \|\gamma\|_{V^r} + O(\|\gamma\|_{V^r}^2)$$

whenever  $\|\gamma\|_{V^r} \leq \delta$ . (We allow the  $O()$  constants here to depend on  $r$ , the Lie group  $G$ , and the norm structure, but not on  $\delta$ ). Let us now see why these estimates will prove the lemma. Let us begin by showing that (76) implies (74). Certainly this will be the case if  $\gamma_l$  has  $V^r$  norm less than  $\delta$ . If instead  $\gamma_l$  has  $V^r$  norm larger than  $\delta$ , we can use Lemma C.2 repeatedly to partition it into  $O(\delta^{-r} \|\gamma_l\|_{V^r}^r)$  curves, all of whose  $V^r$  norms are less than  $\delta$ . These curves are the left-traces of various components of  $\gamma$ , and thus by (76) these components have a  $V^r$  norm bounded by some quantity depending on  $\delta$ .



Concatenating these components together (using the triangle inequality) we obtain the result. A similar argument allows one to deduce (75) from (77).

Next, we observe that to prove the two estimates (76), (77) it suffices to just prove one of the two, for instance (76), as this will also imply (77) for  $\|\gamma\|_{V^r}$  sufficiently small by the usual continuity argument (look at the set of times  $t$  for which the restriction of  $\gamma$  to  $[a, b]$  obeys a suitable version of (77), and use (76) to show that this set is both open and closed if  $\|\gamma\|_{V^r}$  is small enough).

It remains to prove (76) for  $\delta$  sufficiently small. We shall in fact prove the more precise statement (note  $\gamma_l(a) = 0$ )

$$(78) \quad \|\log(\gamma(a)^{-1}\gamma(b)) - \gamma_l(b)\|_{\mathfrak{g}} \leq K\|\gamma_l\|_{V^r}^2$$

for some absolute constant  $K > 0$  (and for  $\delta$  sufficiently small), where  $\log$  is the inverse of the exponential map  $\exp : \mathfrak{g} \rightarrow G$ . Note that it follows from a continuity argument as in the previous paragraph that if  $\delta$  is sufficiently small then  $\gamma(b)^{-1}\gamma(a)$  is sufficiently close to the identity so that the logarithm is well-defined. Let us now see why (78) implies (76). Applying the inequality to any segment  $[t_j, t_{j+1}]$  in  $[a, b]$  we see that

$$\|\log(\gamma(t_j)^{-1}\gamma(t_{j+1})) - (\gamma_l(t_{j+1}) - \gamma_l(t_j))\|_{\mathfrak{g}} \leq K\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r}^2$$

and hence with Lemma C.1 (since  $\delta$  is small)

$$d(\gamma(t_{j+1}), \gamma(t_j)) = \|\gamma_l(t_{j+1}) - \gamma_l(t_j)\|_{\mathfrak{g}} + O(\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r}^2).$$

Estimating  $O(\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r}^2)$  crudely by  $\|\gamma_l\|_{V^r} O(\|\gamma_l|_{[t_j, t_{j+1}]}\|_{V^r})$  and taking the  $\ell^r$  sum in the  $j$  index, we see that for any partition  $a = t_0 < \dots < t_n = b$  we have

$$\left(\sum_{j=0}^{n-1} d(\gamma(t_{j+1}), \gamma(t_j))^r\right)^{1/r} = \left(\sum_{j=0}^{n-1} \|\gamma_l(t_{j+1}) - \gamma_l(t_j)\|_{\mathfrak{g}}^r\right)^{1/r} + O(\|\gamma_l\|_{V^r}^2).$$

Taking suprema over all partitions we obtain the result.

It remains to prove (78) for some suitably large  $K$ . This we shall do by an induction on scale (or “Bellman function”) argument. Let us fix the smooth curve  $\gamma$ . We shall prove the estimate for all subcurves of  $\gamma$ , i.e. for all intervals  $[t_1, t_2]$  in  $[a, b]$ , we shall prove that

$$(79) \quad \|\log(\gamma(t_1)^{-1}\gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_1))\|_{\mathfrak{g}} \leq K\|\gamma_l|_{[t_1, t_2]}\|_{V^r}^2.$$

Let us first prove this in the case when the interval  $[t_1, t_2]$  is sufficiently short, say of length at most  $\epsilon$  for some very small  $\epsilon$  (depending on  $\gamma$ ). In that case, we perform a Taylor expansion to obtain

$$(80) \quad \gamma_l(t) = \gamma_l(t_1) + \gamma_l'(t_1)(t - t_1) + \frac{1}{2}\gamma_l''(t_1)(t - t_1)^2 + O_\gamma((t - t_1)^3)$$

and

$$(81) \quad \gamma_l'(t) = \gamma_l'(t_1) + \gamma_l''(t_1)(t - t_1) + O_\gamma((t - t_1)^2)$$

when  $t \in [t_1, t_2]$ , and where the  $\gamma$  subscript in  $O_\gamma$  means that the constants here are allowed to depend on  $\gamma$  (more specifically, on the  $C^3$  norm of  $\gamma$ ), and the  $O()$  is with respect to the  $\|\cdot\|_{\mathfrak{g}}$  norm. Also we remark that as  $\gamma$  is assumed smooth,  $\gamma_l'(t_1)$  is bounded away from zero. It is then an easy matter to conclude that

$$(82) \quad \|\gamma_l|_{[t_1, t_2]}\|_{V^r} \geq \frac{1}{2}\|\gamma_l'(t_1)\|_{\mathfrak{g}}|t_2 - t_1|$$

if  $\epsilon$  is sufficiently small depending on  $\gamma$ . On the other hand, from (72) and (81) we have

$$\gamma'(t) = \gamma(t)(\gamma'_l(t_1) + \gamma''_l(t_1)(t - t_1) + O_\gamma((t - t_1)^2))$$

from which one may conclude that

$$\gamma(t) = \gamma(t_1) \exp(\gamma'_l(t_1)(t - t_1) + \frac{1}{2}\gamma''_l(t_1)(t - t_1)^2 + O(\|\gamma'_l(t_1)\|_{\mathfrak{g}}|t - t_1|^2) + O_\gamma((t - t_1)^3))$$

for all  $t \in [t_1, t_2]$ , if  $\gamma$  is sufficiently small. We rewrite this as

$$\begin{aligned} \log(\gamma(t_1)^{-1}\gamma(t)) \\ = \gamma'_l(t_1)(t - t_1) + \frac{1}{2}\gamma''_l(t_1)(t - t_1)^2 + O(\|\gamma'_l(t_1)\|_{\mathfrak{g}}|t - t_1|^2) + O_\gamma((t - t_1)^3), \end{aligned}$$

and then specialize to the case  $t = t_2$ . By (80), we have

$$\log(\gamma(t_1)^{-1}\gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_1)) = O(\|\gamma'_l(t_1)\|_{\mathfrak{g}}|t_2 - t_1|^2) + O_\gamma((t_2 - t_1)^3),$$

and hence by (82) we have (79) if  $t_2 - t_1$  is small enough (depending on  $\gamma$ ) and  $K$  is large enough (*independent* of  $\gamma$ ).

This proves (79) when the interval  $[t_1, t_2]$  is small enough. By (82), it also proves (79) when  $\|\gamma_l|_{[t_1, t_2]}\|_{V^r}$  is sufficiently small. To conclude the proof of (79) in general, we now assert the following inductive claim: if (79) holds whenever  $\|\gamma_l|_{[t_1, t_2]}\|_{V^r} < \epsilon$  and some given  $0 < \epsilon \leq \delta$ , then it also holds whenever  $\|\gamma_l|_{[t_1, t_2]}\|_{V^r} < 2^{1/r}\epsilon$ , providing that  $K$  is sufficiently large (*independent* of  $\epsilon$ ) and  $\delta$  is sufficiently small (depending on  $K$ , but *independent* of  $\epsilon$ ). Iterating this we will obtain the claim (79) for all intervals  $[t_1, t_2]$  in  $[a, b]$ .

It remains to prove the inductive claim. Let  $[t_1, t_2]$  be any subinterval of  $[a, b]$  such that the quantity  $A = \|\gamma_l|_{[t_1, t_2]}\|_{V^r}$  is less than  $2^{1/r}\epsilon$ . Applying Lemma C.2, we may subdivide  $[t_1, t_2] = [t_1, t_*] \cup [t_*, t_2]$  such that

$$\|\gamma_l|_{[t_1, t_*]}\|_{V^r}, \|\gamma_l|_{[t_*, t_2]}\|_{V^r} \leq 2^{-1/r}A < \epsilon \leq r.$$

By the inductive hypothesis, we thus have

$$\|\log(\gamma(t_1)^{-1}\gamma(t_*)) - (\gamma_l(t_*) - \gamma_l(t_1))\|_{\mathfrak{g}} \leq K2^{-2/r}A^2$$

and

$$\|\log(\gamma(t_*)^{-1}\gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_*))\|_{\mathfrak{g}} \leq K2^{-2/r}A^2.$$

In particular, we have

$$\begin{aligned} \|\log(\gamma(t_1)^{-1}\gamma(t_*))\|_{\mathfrak{g}} &\leq \|\gamma_l(t_*) - \gamma_l(t_1)\|_{\mathfrak{g}} + K2^{-2/r}A^2 \\ &\leq \|\gamma_l|_{[t_1, t_*]}\|_{V^r} + O(KA^2) \\ &= O(A(1 + KA)) = O(A(1 + K\delta)) = O(A) \end{aligned}$$

if  $\delta$  is sufficiently small depending on  $K$ . Similarly we have

$$\|\log(\gamma(t_*)^{-1}\gamma(t_2))\|_{\mathfrak{g}} = O(A)$$

and hence by the Baker-Campbell-Hausdorff formula (if  $\delta$  is sufficiently small)

$$\|\log(\gamma(t_1)^{-1}\gamma(t_2)) - \log(\gamma(t_1)^{-1}\gamma(t_*)) - \log(\gamma(t_*)^{-1}\gamma(t_2))\|_{\mathfrak{g}} = O(A^2).$$

By the triangle inequality, we thus have

$$\|\log(\gamma(t_1)^{-1}\gamma(t_2)) - (\gamma_l(t_2) - \gamma_l(t_1))\|_{\mathfrak{g}} \leq 2K2^{-2/r}A^2 + O(A^2).$$

We now use the hypothesis  $r < 2$ , which forces  $2 \times 2^{-2/r} < 1$ . If  $K$  is large enough (depending on  $r$ , but independently of  $\delta$ ,  $A$ , or  $\epsilon$ ) we thus have (79). This closes the inductive argument.  $\square$

Letting  $w, v$  be any elements of the Lie algebra  $\mathfrak{g}$ , one can define a nonlinear Fourier summation operator associated to  $G, w, v$  by means of the left trace

$$\begin{aligned}\mathcal{NC}[f](k, 0) &= I \\ \frac{\partial}{\partial x} \mathcal{NC}[f](k, x) &= \mathcal{NC}[f](k, x) \left( \operatorname{Re}(e^{-2\pi i k x} f(x)) w + \operatorname{Im}(e^{-2\pi i k x} f(x)) v \right)\end{aligned}$$

or (giving a different operator) by the right trace

$$\begin{aligned}\mathcal{NC}[f](k, 0) &= I \\ \frac{\partial}{\partial x} \mathcal{NC}[f](k, x) &= \left( \operatorname{Re}(e^{-2\pi i k x} f(x)) w + \operatorname{Im}(e^{-2\pi i k x} f(x)) v \right) \mathcal{NC}[f](k, x).\end{aligned}$$

Above,  $k, x \in \mathbb{R}$ ,  $\mathcal{NC}[f]$  takes values in  $G$ ,  $I$  is the identity element of  $G$ , and  $\operatorname{Re}, \operatorname{Im}$  are the real and imaginary parts of a complex number. An example of interest is given by  $G = SU(1, 1)$ , and

$$w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad v = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

Combining Lemma C.3 with the variational Menshov-Paley-Zygmund theorem of the previous section, we obtain a variational version of the Christ-Kiselev theorem [6]. Namely, we see that for  $1 \leq p < 2$  and  $r > p$

$$\left\| \mathbb{1}_{|\mathcal{NC}[f]| \leq 1} \mathcal{NC}[f] \right\|_{L_k^{p'}(V_x^r)} \leq C_{p,r,G,w,v} \|f\|_{L^p(\mathbb{R})}$$

and

$$\left\| \mathbb{1}_{|\mathcal{NC}[f]| \geq 1} \mathcal{NC}[f] \right\|_{L_k^{p'/r}(V_x^r)}^{1/r} \leq C_{p,r,G,w,v} \|f\|_{L^p(\mathbb{R})}.$$

Note that the usual logarithms are hidden in the  $d$  metric we have placed on the Lie group  $G$ .

Extending these estimates to the case  $p = 2$  is an interesting and challenging problem, even when  $r = \infty$ , which would correspond to a nonlinear Carleson theorem. Lemma C.3 cannot be extended to any exponent  $r \geq 2$ . Sandy Davie and the fifth author of this paper have an unpublished example of a curve in the Lie group  $SU(1, 1)$  with trace in the subspace of  $\mathfrak{su}(1, 1)$  of matrices vanishing on the diagonal so that the diameter of the curve is not controlled by the 2-variation of the trace.

Terry Lyons' machinery [24] via iterated integrals faces an obstruction in a potential application to a nonlinear Carleson theorem because of the unboundedness results for the iterated integrals shown in [26].

#### D. AN APPLICATION TO ERGODIC THEORY

Wiener-Wintner type theorems is an area in ergodic theory that is most closely related to the study of Carleson's operator. In [19], Lacey and Terwilleger prove the following singular integral variant of the Wiener-Wintner theorem:

**Theorem D.1.** *For  $1 < p$ , all measure preserving flows  $\{T_t : t \in \mathbb{R}\}$  on a probability space  $(X, \mu)$  and functions  $f \in L^p(\mu)$ , there is a set  $X_f \subset X$  of probability one, so that for all  $x \in X_f$  we have that the limit*

$$\lim_{s \rightarrow 0} \int_{s < |t| < 1/s} e^{i\theta t} f(T_t x) \frac{dt}{t} .$$

*exists for all  $\theta \in \mathbb{R}$ .*

One idea to approach such convergence results is to study quantitative estimates in the parameter  $s$  that imply convergence, as pioneered by Bourgain's paper [1] in similar context. We first need to pass to a mollified variant of the above theorem:

**Theorem D.2.** *Let  $\phi$  be a function on  $\mathbb{R}$  in the Wiener space, i.e. the Fourier transform  $\widehat{\phi}$  is in  $L^1(\mathbb{R})$ . For  $1 < p$ , all measure preserving flows  $\{T_t : t \in \mathbb{R}\}$  on a probability space  $(X, \mu)$  and functions  $f \in L^p(\mu)$ , there is a set  $X_f \subset X$  of probability one, so that for all  $x \in X_f$  we have that the limits*

$$\begin{aligned} \lim_{s \rightarrow \infty} \int e^{i\theta t} f(T_t x) \phi(st) \frac{dt}{t} , \\ \lim_{s \rightarrow 0} \int e^{i\theta t} f(T_t x) \phi(st) \frac{dt}{t} . \end{aligned}$$

*exist for all  $\theta \in \mathbb{R}$ .*

This theorem clearly follows from an a priori estimate

$$\left\| \sup_{\theta} \left\| \int e^{i\theta t} f(T_t x) \phi(st) \frac{dt}{t} \right\|_{V^r(s)} \right\|_{L^p(x)} \leq C \|f\|_{L^p}$$

for  $r > \max(2, p')$ . Here we have written  $V^r(s)$  for the variation norm taken in the parameter  $s$  of the expression inside, and likewise for  $L^p(x)$ . The variation norm is the strongest norm widely used in this context, while Lacey and Terwilliger use a weaker oscillation norm in the proof of their Theorem.

By a standard transfer method ([3], [7]) involving replacing  $f$  by translates  $T_y f$  and an averaging procedure in  $y$ , the a priori estimate can be deduced from an analogous estimate on the real line

$$(83) \quad \left\| \sup_{\xi} \left\| \int e^{\xi i t} f(x+t) \phi(st) \frac{dt}{t} \right\|_{V^r(s)} \right\|_{L^p(x)} \lesssim \|f\|_{L^p} .$$

The main purpose of this appendix is to show how this estimate (83) can be deduced from the main theorem of this paper by an averaging argument. We write the  $V^r(s)$  norm explicitly and expand  $\phi$  into a Fourier integral to obtain for the left hand side of (83) the expression

$$\left\| \sup_{\xi} \sup_{s_0 < s_1 < \dots < s_K} \left( \sum_{k=1}^K \left| \int \int e^{\xi i t} f(x+t) e^{i\eta(s_k - s_{k-1})t} \frac{dt}{t} \widehat{\phi}(\eta) d\eta \right|^r \right)^{1/r} \right\|_{L^p(x)} .$$

Now pulling the integral in  $\eta$  out of the various norms and considering only positive  $\eta$  (with the case of negative  $\eta$  being similar) and defining  $\xi_k = \xi + \eta s_k$  we obtain the upper

bound

$$\int_{\eta>0} \left\| \sup_{\xi_0<\xi_1<\dots<\xi_K} \left( \sum_{k=1}^K \left| \int e^{i(\xi_k-\xi_{k-1})t} f(x+t) \frac{dt}{t} \right|^r \right)^{1/r} \right\|_{L^p(x)} |\widehat{\phi}(\eta)| d\eta .$$

Now applying the variational Carleson estimate and doing the trivial integral in  $\eta$  bounds this term by a constant times  $\|f\|_{L^p}$ .

*Remark D.3.* To prove the Lacey-Terwiler theorem D.1 from the mollified version, one may approximate the characteristic functions used as cutoff functions by Wiener space functions so that the difference is small in  $L^1$  norm. Then at least for  $f$  in  $L^\infty$  one can show convergence of the limits by an approximation argument, even though one will not recover the full strength of the quantitative estimate in the Wiener space setting. The result for  $f$  in  $L^\infty$  can then be used as a dense subclass result in other  $L^p$  spaces, which can be handled by easier maximal function estimates and further approximation arguments.

*Remark D.4.* The classical version of the Wiener-Wintner theorem does not invoke singular integrals but more classical averages of the type

$$\frac{1}{2s} \int_{|t|<s} e^{i\theta t} f(T_t x) dt .$$

We note that the same technique as above may be applied to these easier averages.

## REFERENCES

- [1] Jean Bourgain, *Pointwise ergodic theorems for arithmetic sets*, Inst. Hautes Études Sci. Publ. Math. (1989), no. 69, 5–45. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.
- [2] James T. Campbell, Roger L. Jones, Karin Reinhold and Máté Wierdl, *Oscillation and variation for the Hilbert transform*, Duke Math. J., **105** (2000), 59–83.
- [3] Alberto P. Calderón, *Ergodic theory and translation-invariant operators*, Proc. Nat. Acad. Sci. U.S.A. **59** (1968) 349–353.
- [4] Lennart Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.
- [5] Michael Christ and Alexander Kiselev, *Maximal functions associated to filtrations*, J. Funct. Anal. **179** (2001), no. 2, 409–425.
- [6] ———, *WKB asymptotic behavior of almost all generalized eigenfunctions for one-dimensional Schrödinger operators with slowly decaying potentials*, J. Funct. Anal. **179** (2001), no. 2, 426–447.
- [7] Ronald R. Coifman and Guido Weiss, *Transference methods in analysis*. Conference Board of the Mathematical Sciences Regional Conference Series in Mathematics, No. 31. American Mathematical Society, Providence, R.I., 1976.
- [8] Ciprian Demeter, Michael T. Lacey, Terence Tao, and Christoph Thiele, *Breaking the duality in the return times theorem*, Duke Math. J. **143** (2008), no. 2, 281–355.
- [9] ———, *The Walsh model for  $M_2^*$  Carleson*, Rev. Mat. Iberoamericana **24** (2008), no. 3, 721–744.
- [10] Charles Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. (2) **98** (1973), 551–571.
- [11] Charles Fefferman and Elias M. Stein,  *$H^p$  spaces of several variables*. Acta Math. **129** (1972), no. 3–4, 137–193.
- [12] Loukas Grafakos, Terence Tao, and Erin Terwiler,  *$L^p$  bounds for a maximal dyadic sum operator*, Math. Z. **246** (2004), no. 1–2, 321–337.
- [13] Richard A. Hunt, *On the convergence of Fourier series*, Orthogonal Expansions and their Continuous Analogues (Proc. Conf., Edwardsville, Ill., 1967), Southern Illinois Univ. Press, Carbondale, Ill., 1968, pp. 235–255.
- [14] Roger L. Jones, Robert Kaufman, Joseph Rosenblatt, and Máté Wierdl, *Oscillation in ergodic theory*, Erg. Th. & Dyn. Sys., **18** (1998), 889–936.

- [15] Roger L. Jones, Andreas Seeger, and James Wright, *Strong variational and jump inequalities in harmonic analysis*, Trans. Amer. Math. Soc. **360** (2008), no. 12, 6711–6742.
- [16] Roger L. Jones and Gang Wang, *Variation inequalities for the Fejér and Poisson kernels*, Trans. Amer. Math. Soc. **356** (2004), no. 11, 4493–4518.
- [17] Carlos E. Kenig and Peter A. Tomas, *Maximal operators defined by Fourier multipliers*, Studia Math. **68** (1980), no. 1, 79–83.
- [18] Michael Lacey, *Issues related to Rubio de Francia's Littlewood-Paley inequality*, NYJM Monographs, **2**. State University of New York, University at Albany, Albany, NY, (2007) 36 pp. (electronic).
- [19] Michael Lacey and Erin Terwilleger, *A Wiener-Wintner theorem for the Hilbert transform*, Ark. Mat. **46** (2008), no. 2, 315–336.
- [20] Michael Lacey and Christoph Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. **7** (2000), no. 4, 361–370.
- [21] Karel de Leeuw, *On  $L_p$  multipliers*, Ann. of Math. **81** (1965), 364–379.
- [22] Dominique Lépingle, *La variation d'ordre  $p$  des semi-martingales*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **36** (1976), no. 4, 295–316.
- [23] Terry Lyons, *Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young*, Math. Res. Lett. **1** (1994), no. 4, 451–464.
- [24] ———, *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana **14** (1998), 215–310.
- [25] Camil Muscalu, Terence Tao, and Christoph Thiele, *A Carleson theorem for a Cantor group model of the scattering transform*, Nonlinearity **16** (2003), no. 1, 219–246.
- [26] ———, *A counterexample to a multilinear endpoint question of Christ and Kiselev*, Math. Res. Lett. **10** (2003) no. 2-3, 237–246.
- [27] Gilles Pisier and Quanhua Xu, *The strong  $p$ -variation of martingales and orthogonal series*, Prob. Theory Related Fields, **77** (1988), 497–451.
- [28] Gabriel Teodor Pripoe, *Vector fields dynamics as geodesic motion on Lie groups* C. R. Math. Acad. Sci. Paris **342** (2006), no. 11, 865–868.
- [29] Jinghua Qian, *The  $p$ -variation of partial sum processes and the empirical process*, Ann. Probab. **26** (1998), no. 3, 1370–1383.
- [30] Tong Seng Quek, *Littlewood-Paley type inequality on  $\mathbb{R}$* , Math. Nachr. 248/249 (2003), 151–157.
- [31] José L. Rubio de Francia, *A Littlewood-Paley inequality for arbitrary intervals*, Revista Mat. Iberoamericana **1** (1985), no. 2, 1–14.
- [32] Elias M. Stein and Guido Weiss, *Introduction to Fourier analysis on Euclidean spaces*. Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971.
- [33] Terence Tao and Christoph Thiele, *Nonlinear Fourier analysis*, IAS/Park City Math. Ser., to appear.
- [34] Christoph Thiele, *Wave packet analysis*, CBMS Regional Conference Series in Mathematics, vol. 105, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2006.
- [35] Norbert Wiener, *The quadratic variation of a function and its Fourier coefficients*, MIT Journal of Math. and Physics **3**(1924) pp. 72–94.

R. OBERLIN, DEPARTMENT OF MATHEMATICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, LA 70803-4918, USA

*E-mail address:* richard.oberlin@gmail.com

A. SEEGER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN, 480 LINCOLN DRIVE, MADISON, WI, 53706, USA

*E-mail address:* seeger@math.wisc.edu

T. TAO, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095-1555, USA

*E-mail address:* tao@math.ucla.edu

C. THIELE, DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095-1555, USA

*E-mail address:* `thiele@math.ucla.edu`

J. WRIGHT, MAXWELL INSTITUTE OF MATHEMATICAL SCIENCES AND THE SCHOOL OF MATHEMATICS,  
UNIVERSITY OF EDINBURGH, JCMB, KING'S BUILDINGS, MAYFIELD ROAD, EDINBURGH EH9 3JZ,  
SCOTLAND

*E-mail address:* `J.R.Wright@ed.ac.uk`